REVIEW

A gauge-theoretic approach to gravity

BY KIRILL KRASNOV[1,2,*]

[1]*School of Mathematical Sciences, University of Nottingham, University Park, Nottingham NG7 2RD, UK*
[2]*Max Planck Institute for Gravitational Physics (Albert Einstein Institute), Am Mühlenberg 1, 14476 Golm, Germany*

Einstein's general relativity (GR) is a dynamical theory of the space–time metric. We describe an approach in which GR becomes an SU(2) gauge theory. We start at the linearized level and show how a gauge-theoretic Lagrangian for non-interacting massless spin two particles (gravitons) takes a much more simple and compact form than in the standard metric description. Moreover, in contrast to the GR situation, the gauge theory Lagrangian is convex. We then proceed with a formulation of the full nonlinear theory. The equivalence to the metric-based GR holds only at the level of solutions of the field equations, that is, on-shell. The gauge-theoretic approach also makes it clear that GR is not the only interacting theory of massless spin two particles, in spite of the GR uniqueness theorems available in the metric description. Thus, there is an infinite-parameter class of gravity theories all describing just two propagating polarizations of the graviton. We describe how matter can be coupled to gravity in this formulation and, in particular, how both the gravity and Yang–Mills arise as sectors of a general diffeomorphism-invariant gauge theory. We finish by outlining a possible scenario of the ultraviolet completion of quantum gravity within this approach.

> Since the mathematicians have invaded the theory of relativity, I do not understand it myself anymore.
>
> Albert Einstein

1. Introduction

This expository paper is about an approach to the theory of general relativity (GR) that departs rather far from how its initiator Albert Einstein thought about the subject. The present author is a physicist by training but working in a mathematics department, and so it is quite possible that Einstein would be able to say about this work what he said on a different occasion and what is quoted above. The approach we shall describe makes a conceptual jump when compared with Einstein's views on the nature of the gravitational force. Thus, in

*[*kirill.krasnov@nottingham.ac.uk](mailto:kirill.krasnov@nottingham.ac.uk)

Einstein's GR, the gravitational field is encoded in the geometrical properties of the space–time, and the geometry itself is described by the space–time metric. The approach described here will still be geometrical in spirit. However, the geometry will be that of *gauge fields* on the space–time manifold, not that of a space–time metric. No metric will be present in the formulation of the theory, and it will only appear later as a derived concept.

Accepting for the moment that such an approach is indeed possible, a natural question to ask is if it is worth considering any such reformulation of GR. Indeed, it is clear that it would require drastic changes to the century-old metric-based intuition about the subject. Such conceptual changes can only be justified if they can help to solve problems that are difficult to address in the traditional framework. As we now briefly review, there are indeed at least two major difficulties with GR.

The first problem of GR is that of the unification with the other forces. Thus, Einstein's theory was developed already almost 100 years ago, in particular, before the theory of quantum mechanics was discovered. And during the last century, our understanding of the behaviour of the micro-world has culminated in the discovery that almost all interactions in Nature (apart from gravity) are described by (quantum) gauge fields. Thus, it is now believed that (to the energy scales of the conjectural grand unification) the electromagnetic, the weak and the strong interactions are all described by the Yang–Mills theory of the (spontaneously broken) Standard Model gauge group $U(1) \times SU(2) \times SU(3)$. In other words, at least to a very high energy scale, the gauge fields appear to be fundamental. Only the gravitational force stands far aside from this compelling pattern. Using the technical jargon, we say that gauge fields describe particles of spin 1, while the gravitational interactions are carried by particles of spin 2, and this is where the difference comes from. At the same time, if it was possible to reformulate the theory of gravity in the language closer to that of gauge theories, we would possibly come closer to understanding how gravity fits together with the other forces of Nature. To put it differently, it is not clear that Einstein himself would be thinking in terms of the space–time metric if he was trying to invent his theory of gravity nowadays, with our present solid appreciation of the importance of the gauge principle.

The other problem of GR is related to the quantum properties of the gravitational field. These do not seem to be of any importance at energy scales accessible to our observations or experiments. At the same time, they can be studied theoretically and reveal that gravity is again unlike all the other forces. For other forces, the quantum vacuum polarization effects lead to the strength of interactions becoming energy scale-dependent (the phenomenon of renormalization). Yet, the form of the field equations (or the Lagrangian) describing these interactions does not change with the energy, only the coupling constants (and other physical parameters, e.g. masses) flow with the scale. One says that the theories that form the standard model of elementary particles are renormalizable. Gravity is not like this. Not only the parameters of its Lagrangian flow with the energy, but also the form of the Lagrangian itself must be modified once the quantum corrections are taken into account. One says that gravity is non-renormalizable. What this means is that we know for certain that Einstein's theory of GR cannot be a fundamental theory, for at very high energies, it is modified (by quantum corrections) in a way we cannot currently control.

A possible way out of the abovementioned difficulties is to imbed both gravity and the gauge-fields in a different theory that resolves these problems. This is, for example, what string theory attempts to do. However, the string theory, at least in the way we currently understand it, appears to be far from being an economical solution of the problem, as it requires supersymmetry as well as many extra unseen dimensions. It is not impossible that a more economical solution to the problems of gravity could exist.

As we shall explain in this article, the present gauge theory approach may be of help with both the problem of unification with other interactions as well as the problem of the quantum behaviour. We emphasize, however, that neither of these two goals has been achieved as of yet, there are only hints (to be described) that this may be possible. A sceptically minded reader may then conclude that in the absence (as of yet) of any real applications, the approach presented here is not worth the trouble learning, for it requires developing a completely new intuition (gauge field-based) about the gravitational force.

At the same time, we note that in the past, science has often progressed by first reformulating a known physical law in a non-trivially equivalent form, and then this non-trivial reformulation suggesting a generalization leading to a new physical theory. For this author, the most striking example of this historical phenomenon is the discovery of quantum mechanics. Indeed, at least in its first formulations, it was so essentially built on the Hamiltonian formulation of the classical physics. Thus, it appears to be profitable to reformulate known physical laws in a form that is non-trivially equivalent to the original formulation. This way we not only learn more about the theory at hand, but also potentially step on the path to new physics. We therefore proceed with our development of a gauge-theoretic description of gravity having at least this historic lesson as the motivation.

As for a mathematically minded reader, we motivate our constructions by the following remarks. In spite of Yang–Mills theory and gravity being so different in spirit—the former is about connections in principal G-bundles while the latter is about metric-compatible connections in the tangent bundle—a subtle interplay between the two theories has become apparent via work on Yang–Mills instantons in the late-1970s. This is particularly clearly illustrated by the fact that the index 1 SU(2) instantons (self-dual solutions of the Yang–Mills field equations) are all gauge-equivalent to the self-dual part of the gravitational Levi-Civita connection for metrics conformally related to the standard round metric on S^4 (see Atiyah *et al.* (1978), theorem 9.1). Rephrasing, one can say that all index 1 SU(2) instantons come from gravitational instantons. Thus, at least in the self-dual case, there is an intimate relation between the solutions of the two theories. It is this relation that is going to be at the root of the developments reviewed in this paper, with the notion of the self-duality playing the central role.

The final word of caution is as follows. We make no claim here that by simply reformulating gravity in the language of gauge fields, one comes closer to understanding the problems of quantum gravity and unification. In fact, such reformulations were around for a long time, with historically the first being those due to Plebanski (1977) and MacDowell & Mansouri (1977), then followed by a new Hamiltonian formulation owing to Ashtekar (1987). These formulations of GR do shed some new light on the problems of gravity, but do not by themselves provide a solution of these problems. So, the novelty of the approach

described in this paper is not in the fact that a gauge-theoretic reformulation of gravity is possible—this was known for quite some time. It is in the fact that the specific gauge theory description to be reviewed here has certain attractive features not shared by other formulations, and these may eventually help us to come to terms with the abovementioned problems of gravity. Thus, one of the most interesting features of the new formulation is that the (Euclidean signature) action functional becomes convex, a very desirable property that the usual metric-based formulation does not have. This feature is in turn related to other simplifications in the structure of the theory, which is what gives us hope that the new formulation can shed new light on the problems of gravity.

With these introductory remarks being made, we shall proceed with our gauge-theoretic description of gravity. We shall follow the bottom-to-top approach, in which we will first present a gauge theory description of gravity in as least general terms as possible, and then proceed to exhibit the general principle. Thus, in §2, we show how the linearized theory (i.e. gravitons) can be described in the gauge theory terms. We present the construction of the full nonlinear theory in §3. This section also discusses how the unification is realized in the present approach. We then discuss the quantum theory in §4. We reiterate the main points of our approach in §5.

2. Gravity as a gauge theory: linearized level

We start by showing how gravity can be described in gauge-theoretic terms at the level of the linearized theory. This is already non-trivial, because the statement that gauge fields are spin 1, while gravitons are spin 2 particles is the statement about the corresponding linearized theories.

From the outset, we note the fact that the linearized GR described in terms of connections is not new. For instance, this is possible in the framework of Plebanski formulation of GR (Plebanski 1977) and related to it (via the space plus time decomposition) Ashtekar Hamiltonian formulation (Ashtekar 1987). The principal difference between these and our approach is that our linearized theory Lagrangian will be a functional of only the connection field, with no other (auxiliary) fields necessary (as in, e.g. Plebanski 1977). We shall see that this leads to some important simplifications, and that our linearized theory will be arguably simpler than all other known descriptions of the linearized gravity. In particular, as we shall see below, the diffeomorphisms are realized in our gauge-theoretic description in a particularly simple way (different from all other existing descriptions), and the corresponding gauge components of the field can be projected away from the very beginning, leading to considerable simplifications in the structure of the theory. In addition, we shall see that our Lagrangian for the linearized GR has very much the same form as that for the linearized Yang–Mills theory, which will eventually be of help when we put the two theories together.

The gauge-theoretic description of free massless spin 2 particles given in this section appears to be new, and was first given by Krasnov (2011*a*).

(*a*) *Linearized Yang–Mills theory*

Let us start our description by briefly reviewing the linearization of the Yang–Mills theory. Let A be a connection in the principal G-bundle over the space–time

M, and let $F = \mathrm{d}A + (1/2)[A, A]$ be its curvature two-form. Here, we assume G to be compact. We use the conventions in which $F = (1/2)F_{\mu\nu}\,\mathrm{d}x^{\mu} \wedge \mathrm{d}x^{\nu}$ and often suppress the Lie algebra index of F for brevity. If $g_{\mu\nu}$ is a (given) space–time metric, which we for now assume to be that of the Minkowski space–time, then it can be used to raise the indices of $F_{\mu\nu}$ to get $F^{\mu\nu}$, which can then be contracted with $F_{\mu\nu}$ to form the familiar Yang–Mills Lagrangian

$$\mathcal{L}_{\mathrm{YM}} = -\frac{1}{4g_{YM}^2}\mathrm{Tr}(F^{\mu\nu}F_{\mu\nu}), \tag{2.1}$$

where the trace stands for a (in our conventions positive-definite) Killing–Cartan form on the Lie algebra of G, appropriately normalized. Thus, the trace can also be replaced by a sum over the Lie algebra indices of the curvature squared. At the linearized level (expanding around the zero background connection and introducing the canonically normalized perturbation a so that $A = g_{\mathrm{YM}}\,a$), we get

$$\mathcal{L}_{\mathrm{YM}}^{(2)} = -\tfrac{1}{2}\mathrm{Tr}(\partial^{\mu}a^{\nu}(\partial_{\mu}a_{\nu} - \partial_{\nu}a_{\mu})). \tag{2.2}$$

As usual, the need for the minus sign in front of the action is dictated by the desire to have a positive-definite Hamiltonian. Indeed, introducing the space+time split $\mu = (0, i), i = 1, 2, 3$ and taking (for definiteness) the metric signature to be $(-, +, +, +)$ we get

$$\mathcal{L}_{\mathrm{YM}}^{(2)} = \tfrac{1}{2}\mathrm{Tr}((\dot{a}_i - \partial_i a_0)^2 - (B_i)^2), \tag{2.3}$$

where the dot denotes the time derivative and we have introduced a notation $B_i = \epsilon_{ijk}\partial_j a_k$. We therefore see that the momentum canonically conjugate to a_i is $E_i := \dot{a}_i - \partial_i a_0$, and the momentum conjugate to the time component of the connection a_0 is zero. Thus, the time component of the connection plays the role of the Lagrange multiplier imposing the so-called Gauss constraint, and the linearized Hamiltonian $\mathcal{H}_{\mathrm{YM}}^{(2)} = \mathrm{Tr}(E_i\dot{a}_i) - \mathcal{L}_{\mathrm{YM}}^{(2)}$ is given by:

$$\mathcal{H}_{\mathrm{YM}}^{(2)} = \tfrac{1}{2}\mathrm{Tr}((E_i)^2 + (B_i)^2) - \mathrm{Tr}(a_0\partial_i E_i), \tag{2.4}$$

where we have integrated by parts in the last term. The Hamiltonian is explicitly positive-definite (if the gauge group G is compact), modulo a constraint term that imposes the electric field to be trace-free $\partial_i E_i = 0$ and generates the usual gauge transformations of the connection $\delta_{\phi}a_i = \partial_i\phi$, ϕ being an arbitrary Lie algebra-valued function. The Hamiltonian is easily seen to be gauge-invariant, as the electric field does not transform (at this linearized level) and the linearized connection transforms so that the magnetic field B_i is gauge-invariant.

(b) Self-duality

In preparation to the developments in the case of gravity, we will add to the Lagrangian (2.1), a certain total derivative term. This will have no effect on the physics, but will allow us to write the Yang–Mills Lagrangian in a form later almost matched by the gravitational one. To write down the total derivative term required, we need to introduce some new object that will play a very important role later. These objects are certain self-dual two-forms and thus build on the notion of the self-duality on two-forms (in four space–time dimensions that we work in).

Recall that, given a metric $g_{\mu\nu}$ (such as, e.g. the Minkowski metric we have been working with up to now) one has the volume form $\epsilon_{\mu\nu\rho\sigma}$ for $g_{\mu\nu}$. Two of its indices can be raised with the metric to obtain an object $\epsilon_{\mu\nu}{}^{\rho\sigma}$ that can be applied to a two-form $U_{\mu\nu}$ with the result being again a two-form—the Hodge dual of U. We have:

$$U^*_{\mu\nu} := \tfrac{1}{2}\epsilon_{\mu\nu}{}^{\rho\sigma}\,U_{\rho\sigma}. \tag{2.5}$$

As it is not hard to check, in our case of a metric of the Lorentzian signature, the square of the Hodge operator is minus one. Thus, its eigenvalues are $\pm i$, and the space of two-forms splits into two orthogonal subspaces. These are referred to as the spaces of self-dual and anti-self-dual two-forms. We note in passing that these properties of the Hodge duality operator are quire reminiscent of those of a complex structure (on the space of two-forms), with the spaces of self- and anti-self-dual forms being analogous to holomorphic and anti-holomorphic elements. This is more than analogy, and we refer the interested reader to Brans (1974) for a development of this idea.

Thus, for any self-dual two-form $U_{\mu\nu}$ we have:

$$\tfrac{1}{2}\epsilon_{\mu\nu}{}^{\rho\sigma}\,U_{\rho\sigma} = i\,U_{\mu\nu}, \tag{2.6}$$

and for anti-self-dual form, we have an extra minus on the right hand-side (r.h.s.). The space of self-dual two-forms is three-dimensional, and we can introduce a basis in it. A choice of such basic self-dual two-forms can be rather arbitrary as long as they span the required subspace. However, there is always a canonical (modulo certain gauge rotations, see below) *orthonormal* choice of the basis. Let us denote the canonical self-dual two-forms by $\Sigma^i_{\mu\nu}$, $i = 1, 2, 3$. Note that we have denoted the index enumerating the two-forms by the same letter as was used to refer to the spatial index in the Hamiltonian analysis above. This is not an oversight; we shall now see that the two indices can be naturally identified. The canonical self-dual two-forms are required to be orthonormal in that

$$\epsilon^{\mu\nu\rho\sigma}\Sigma^i_{\mu\nu}\Sigma^j_{\rho\sigma} = 8i\delta^{ij}, \tag{2.7}$$

where the numerical coefficient on the r.h.s. is convention-dependent, and δ^{ij} is the Kronecker-delta. It can be shown that the self-dual two-forms satisfying (2.7) are defined uniquely modulo SO(3) rotations preserving δ^{ij}. We can now give an explicit form of the basic self-dual two-forms in the case of the Minkowski space–time metric. Using the two-form notation, we have:

$$\Sigma^i = i\mathrm{d}t \wedge \mathrm{d}x^i + \tfrac{1}{2}\epsilon^{ijk}\,\mathrm{d}x^j \wedge \mathrm{d}x^k, \tag{2.8}$$

it is not hard to check that $\Sigma^i_{\mu\nu}$ are self-dual (with the conventions that $\epsilon^{0123} = +1$), and that (2.7) holds. Let us also note what becomes of the components of the two-forms $\Sigma^i_{\mu\nu}$ under the space+time split. We have:

$$\Sigma^i_{0j} = i\,\delta^i_j \quad \text{and} \quad \Sigma^i_{jk} = \epsilon^i{}_{jk}. \tag{2.9}$$

Thus, we see that the objects $\Sigma^i_{\mu\nu}$ indeed provide a natural identification of the basis index i with the spatial index. This fact will be used later in our discussion of the gravitational case.

For later purposes, let us note an important identity satisfied by our self-dual two-forms. We have

$$\Sigma^{i\,\nu}_{\mu}\Sigma^{j\,\rho}_{\nu} = -\delta^{ij}g_{\mu}{}^{\rho} + \epsilon^{ijk}\Sigma^{k\rho}_{\mu}. \tag{2.10}$$

Thus, the basic self-dual two-forms satisfy an algebra similar to that of Pauli matrices. This identity can be checked by a direct verification, using the explicit expression (2.8).

For our discussion of the Yang–Mills theory, we need one more identity involving $\Sigma^{i}_{\mu\nu}$. Thus, we note that we can use Σ's to construct the projector on the space of self-dual two forms. This is an operator P^{+} whose square coincides with itself, and which projects any two-form onto its self-dual part. We have:

$$P^{+\rho\sigma}_{\mu\nu} \equiv \frac{1}{2}\left(g^{\rho}_{[\mu}g^{\sigma}_{\nu]} - \frac{i}{2}\epsilon_{\mu\nu}{}^{\rho\sigma}\right) = \frac{1}{4}\Sigma^{i}_{\mu\nu}\Sigma^{i\,\rho\sigma}. \tag{2.11}$$

Finally, we note that the self-dual two-forms Σ^{i} have appeared in the literature on many occasions before. In the literature on instantons (self-dual solutions of the Yang–Mills field equations), these objects are often referred to as 't Hooft's symbols ('t Hooft 1976). They will be of fundamental importance in the considerations that follow.

(c) *Linearized Yang–Mills revisited*

We now use the objects introduced in the previous subsection to rewrite the linearised Yang–Mills Lagrangian (plus a surface term) in a convenient form for later purposes. Thus, modulo a total derivative term we have:

$$\mathcal{L}^{(2)}_{\text{YM}} = -2P^{+\mu\nu\rho\sigma}\text{Tr}(\partial_{\mu}a_{\nu}\partial_{\rho}a_{\sigma}). \tag{2.12}$$

Using the basic self-dual two-forms $\Sigma^{i}_{\mu\nu}$ introduced above, we can rewrite this as

$$\mathcal{L}^{(2)}_{\text{YM}} = -\tfrac{1}{2}\text{Tr}(\Sigma^{i\mu\nu}\partial_{\mu}a_{\nu})^{2}. \tag{2.13}$$

An alternative convenient way to write the above Lagrangian is to introduce a basis in the Lie algebra, so that $a_{\mu} = a^{a}_{\mu}T^{a}$, with a being the Lie algebra index and T^{a} being the generators normalized so that $\text{Tr}(T^{a}T^{b}) = \delta^{ab}$. With this in mind, we rewrite the Yang–Mills Lagrangian as

$$\mathcal{L}^{(2)}_{\text{YM}} = -\tfrac{1}{2}\delta_{ab}\delta_{ij}(\Sigma^{i\mu\nu}\partial_{\mu}a^{a}_{\nu})(\Sigma^{j\rho\sigma}\partial_{\rho}a^{b}_{\sigma}). \tag{2.14}$$

Our linearized gravitational Lagrangian below will strongly resemble this form of the Yang–Mills Lagrangian.

An instructive exercise is to repeat the Hamiltonian analysis of Yang–Mills but starting from (2.14). To this end, we need the space+time split of the combination $\Sigma^{i\mu\nu}\partial_{\mu}a^{a}_{\nu}$ that appears prominently in (2.14). We have:

$$\Sigma^{i\mu\nu}\partial_{\mu}a^{a}_{\nu} = -i(\dot{a}^{a}_{i} - \partial_{i}a^{a}_{0}) + \epsilon^{ijk}\partial_{j}a^{a}_{k}. \tag{2.15}$$

We now see that in this formulation of the theory, the momentum conjugate to a_{i} is

$$\pi_{i} := \dot{a}_{i} - \partial_{i}a_{0} + iB_{i} = E_{i} + iB_{i}, \tag{2.16}$$

where E_{i} is the usual electric field and we have used the same definition of the magnetic field B_{i} as above. We have again suppressed the Lie algebra index for

brevity. The Hamiltonian in this formulation reads

$$\mathcal{H}_{\mathrm{YM}}^{(2)} = \mathrm{Tr}(\tfrac{1}{2}(\pi_i)^2 - \mathrm{i}\pi_i B_i) - \mathrm{Tr}(a_0 \partial_i \pi_i). \tag{2.17}$$

This can be rewritten as

$$\mathcal{H}_{\mathrm{YM}}^{(2)} = \tfrac{1}{2}\mathrm{Tr}((\pi_i - \mathrm{i}B_i)^2 + (B_i)^2) - \mathrm{Tr}(a_0 \partial_i \pi_i), \tag{2.18}$$

which is the same Hamiltonian as we have obtained above in (2.4). Indeed, this is obvious for the first term, and for the last Gauss constraint term, the shift of the momentum π_i by a multiple of B_i has no effect as the magnetic field is transverse. We see that in this formulation, the momentum conjugate to the connection is not real, but rather only the combination $\pi_i - \mathrm{i}B_i \equiv E_i$ is real. At the same time, the spatial part a_i of the connection is real.

The reason why the complexity crept in is that we have added to the Lagrangian a term that is a *purely imaginary* total derivative. This has no effect on the dynamics, but it does affect the structure of the Hamiltonian formulation of the theory, in particular, the symplectic structure. This is why the momentum variable is no longer the electric field, but rather the combination (2.16). This corresponds to a canonical transformation on the phase space of Yang–Mills theory with a purely imaginary generating function. The analysis of this subsection is of course very familiar from the discussions of the θ-angle ambiguity in the quantization of Yang–Mills theory (see Ashtekar (1991), Section 13.2) for a very readable account. The only difference with the more familiar case is that the parameter of the canonical transformation is taken to be imaginary. This is certainly a legitimate operation at the classical level, with the price to pay being that one has to be careful about the reality conditions satisfied by the fields. We shall see that similar care will have to be exercised in the gravitational case, where analogous reality conditions are to be imposed.

(d) Linearized gravity

We now turn to the main object of our interest—gravity. We start by simply writing down the linearized Lagrangian of a form similar to (2.14), and later explain how it comes about.

The first step is a choice of the gauge group, and we take $G = \mathrm{SO}(3)$. We shall then denote the Lie algebra index by the same lower case Latin letters from the middle of the alphabet, which we have already used to refer to the spatial indices. Again this is not an oversight, as we shall later see that these types of indices can be naturally identified. We then write a Lagrangian essentially of the same form as (2.14), but with a different tensor used to contract the two quantities $\Sigma^{i\mu\nu}\partial_\mu a_\nu^a$, where a is now to be replaced by an $\mathrm{SO}(3)$ index j. Thus, we write

$$\mathcal{L}_{\mathrm{GR}}^{(2)} = -\tfrac{1}{2} P^{ij|kl}(\Sigma^{i\mu\nu}\partial_\mu a_\nu^j)(\Sigma^{k\rho\sigma}\partial_\rho a_\sigma^l), \tag{2.19}$$

where

$$P^{ij|kl} := \delta^{i(k}\delta^{l)j} - \tfrac{1}{3}\delta^{ij}\delta^{kl} \tag{2.20}$$

is a projector on symmetric trace-free 3×3 matrices. As we shall soon see, the Lagrangian (2.19) describes a spin 2 particle, in contrast to (2.14) that describes a spin 1 particle for each Lie algebra generator. The fact that (2.19) gives a spin 2

particle is related to the fact that (2.20) is a projector on spin 2 representation in the tensor product of two spin 1 representations of SO(3).

Let us now discuss the invariances of the Lagrangian (2.19). As in the case of Yang–Mills theory, (2.19) is obvious invariant under the SO(3) gauge-transformations $\delta_\phi a_\mu^i = \partial_\mu \phi^i$. Indeed, the partial derivatives commute, and the objects $\Sigma^{i\mu\nu}$ are anti-symmetric in their space–time indices. However, unlike in the Yang–Mills case, the Lagrangian (2.19) has an extra invariance, and this is how the diffeomorphisms appear in the game. Consider

$$\delta_\xi a_\mu^i = \xi^\nu \Sigma_{\mu\nu}^i, \tag{2.21}$$

where $\Sigma_{\mu\nu}^i$ are the basic self-dual two-forms (2.8), and ξ^μ is an arbitrary vector field. Note that there are no derivatives present in this transformation rule, this is simply a shift of the linearized connection in a certain direction in the field space. It is not hard to see that our Lagrangian is invariant under such shifts. Indeed, we use the algebra (2.10) of the basic self-dual two-forms and note that the product of two Σ's is in either the spin 0 part (the trace) or in the spin 1 part (the anti-symmetric part) in the tensor product of two spin 1 representations. However, the Lagrangian contains the spin 2 projector (2.20), and so is insensitive to the shifts (2.21). So, (2.21) is an invariance of the Lagrangian. However, as we shall soon see, this is not a gauge invariance in the sense how the usual gauge transformations of the connection were the gauge invariances of the theory in the case of Yang–Mills theory. Thus, we shall see that there are no Lagrange multipliers for the constraints generating (2.21) in the Hamiltonian formulation of the theory. The significance of this fact will be discussed later.

Let us now repeat the exercise of the Hamiltonian analysis of the theory. We have already made most of the steps needed for this in the discussion of the Yang–Mills theory. So, we can simply write the expression for $\Sigma_i^{\mu\nu} \partial_\mu a_{\nu j}$ from (2.15). We have:

$$\Sigma_i^{\mu\nu} \partial_\mu a_{\nu j} = -\mathrm{i}(\dot{a}_{ij} - \partial_i a_{0j}) + \epsilon_i{}^{kl} \partial_k a_{lj} \tag{2.22}$$

where the second index on the spatial connection is the lowered Lie algebra one. In the Lagrangian, this expression is multiplied by the spin 2 projector (2.20). Thus, we immediately see that only the spin 2 part of the connection field a_{ij} is dynamical. The momentum conjugate to this part of the connection is

$$\pi^{ij} = P^{ijkl}((\dot{a}_{kl} - \partial_k a_{0l}) + \mathrm{i}\epsilon_k{}^{mn} \partial_m a_{nl}). \tag{2.23}$$

We see that this expression depends only on the time derivative of the symmetric trace-free part of a_{ij}, on the temporal component of the connection a_{0i}, as well as on spatial derivatives of the full connection a_{ij}. Let us now discuss the dependence on the latter. For this, it is very convenient to split the spatial connection a_{ij} into its irreducible with respect to SO(3) components. Thus, we have:

$$a_{ij} = a_{ij}^{(2)} + a_{ij}^{(1)} + a_{ij}^{(0)}, \tag{2.24}$$

where $a_{ij}^{(2)}$ is the spin 2 part that is symmetric trace-free, $a_{ij}^{(1)}$ is the anti-symmetric spin 1 part and $a_{ij}^{(0)}$ is the spin 0 part proportional to δ_{ij}. We immediately see

that because of the symmetrization imposed by the projector $P^{ij|kl}$, there is no dependence in the last term in (2.23) on the spin 0 part $a_{ij}^{(0)}$. Next, let us write

$$a_{ij}^{(1)} = \epsilon_{ijk} c^k, \tag{2.25}$$

where c^k captures the spin 1 part of the spatial connection. Then, a simple calculation shows that the momentum (2.23) can be written as:

$$\pi^{ij} = P^{ijkl}(\dot{a}_{kl}^{(2)} + \mathrm{i}\epsilon_k{}^{mn}\partial_m a_{nl}^{(2)} - \partial_k(a_{0l} + \mathrm{i}c_l)). \tag{2.26}$$

We thus see that the Lagrangian (2.19) depends only on the spin 2 part of the spatial connection a_{ij}, as well as on the temporal a_{0i} and spin 1 components $a_{ij}^{(1)}$ in the combination $a_{0i} + (\mathrm{i}/2)\epsilon_i{}^{jk} a_{jk}^{(1)}$. There is no dependence at all on the spin 0 part $a_{ij}^{(0)}$.

Let us now find the Hamiltonian $\mathcal{H}_{\mathrm{GR}}^{(2)} = \pi^{ij} a_{ij}^{(2)} - \mathcal{L}_{\mathrm{GR}}^{(2)}$. We have:

$$\mathcal{H}_{\mathrm{GR}}^{(2)} = \tfrac{1}{2}(\pi^{ij})^2 - \mathrm{i}\pi^{ij}\epsilon_i{}^{kl}\partial_k a_{lj} - (a_{0i} + \mathrm{i}c_i)\partial_j \pi^{ij}. \tag{2.27}$$

Here, we have removed the superscript (2) from the spin 2 component of the connection for brevity, with the understanding now that this is the only component of the connection that is dynamical. We have also removed all the projectors $P^{ij|kl}$ because now the momentum π^{ij} is symmetric traceless and so no additional projection is necessary. In addition, we have integrated by parts in the last term.

The structure of (2.27) is already familiar from the discussion of the Yang–Mills theory above. We could write the Hamiltonian in a more suggestive form by introducing the 'magnetic field'

$$B^{ij} := P^{ij|kl}\epsilon_k{}^{mn}\partial_m a_{nl} \tag{2.28}$$

and writing

$$\mathcal{H}_{\mathrm{GR}}^{(2)} = \tfrac{1}{2}((\pi^{ij} - \mathrm{i}B^{ij})^2 + (B^{ij})^2) - (a_{0i} + \mathrm{i}c_i)\partial_j \pi^{ij}. \tag{2.29}$$

This is exactly analogous to (2.18), with the only exception being that the momentum variable π^{ij} as well as the magnetic field B^{ij} are now spin 2 fields. As in the Yang–Mills case, the last term, when varied with respect to the combination $a_{0i} + \mathrm{i}c_i$, generates the constraint that requires the momentum to be transverse. The reality conditions that need to be imposed to guarantee the positive-definiteness of the Hamiltonian are also analogous to the Yang–Mills case. Indeed, we see that the combination $\pi^{ij} - \mathrm{i}B^{ij}$ must be required to be real, as well as the spin 2 part $a_{ij}^{(2)}$ of the spatial connection (and thus the magnetic field B^{ij}).

Thus, we see that at the level of the Hamiltonian formulation (as well as at the level of the action), the treatments of Yang–Mills theory and gravity are exactly analogous. The main difference comes from the fact that in the gravity case, the gauge group has to be taken a specific one SO(3), and that certain propagating modes of the Yang–Mills theory with its two modes per generator have been projected out to leave only two propagating degrees of freedom of gravity. Indeed,

the effect of the projector $P^{ij|kl}$ inserted in the Lagrangian (2.19) is precisely to remove four of the six propagating modes of SO(3) Yang–Mills theory, leaving only two gravitational d.o.f. At the level of the Hamiltonian (2.29), this projection is already carried out because only the spin 2 part of the spatial connection (and its conjugate momentum) is present.

The above discussion of the propagating mode content of (2.19) can be rephrased as follows. The propagating d.o.f. of Yang are described by the spatial connection a_i (with the Lie algebra index suppressed), which can be set (using gauge transformations) to satisfy the transverse condition $\partial^i a_i = 0$. In the case of gravity, we have exactly the same starting point, and the spatial connection a_i^j, where now the Lie algebra index is explicitly indicated. Since the group SO(3) of gauge rotations in this case can be identified with the group of spatial rotations (this identification is provided, e.g. by the spatial projection of the self-dual two-forms $\Sigma_{\mu\nu}^i$), the two indices of a_i^j are on the same footing. Thus, the spatial connection in this case is in the product of two spin 1 representations: $a \in V^1 \otimes V^1$, where our notation is that V^j is the space of irreducible representation of SO(3) of dimension $\dim(V^j) = 2j + 1$. Decomposing $V^1 \otimes V^1$ into irreducibles, we get a direct sum $V^2 \oplus V^1 \oplus V^0$. The projector in (2.19) projects out the spin 1 and spin 0 components, leaving only a single dynamical spin 2 field $a_{ij}^{(2)}$. Gauge transformations (SO(3) rotations) still act on it and can be used to set this field to be transverse. Once this is done, the term $(B^{ij})^2$ becomes $(\partial_i a_{jk})^2$, as can be checked by a simple calculation. Thus, once the gauge has been fixed, we get the following linearized Hamiltonian:

$$\mathcal{H}_{\mathrm{GR}}^{(2)} = \tfrac{1}{2}((p^{ij})^2 + (\partial_i a_{jk})^2), \tag{2.30}$$

where we have denoted the real momentum by $p^{ij} = \pi^{ij} - \mathrm{i}B^{ij}$. This is the usual linearized Hamiltonian of the metric-based GR, as we shall review in §2e.

(e) A comparison to the linearized Einstein–Hilbert action

Above we have seen how a spin 2 particle can be described by a Lagrangian of the Yang–Mills type, but with an additional spin 2 projector inserted. At the level of the Hamiltonian formulation, we have seen the familiar from the metric-based treatment symmetric traceless transverse tensors appearing. The natural question that arises is how the above Lagrangian is related to the usual metric-based GR Lagrangian linearized around the Minkowski metric. The purpose of this subsection is to discuss this relationship.

The Einstein–Hilbert Lagrangian (for our choice of the signature $(-, +, +, +)$) reads:

$$\mathcal{L}_{\mathrm{EH}} = -\frac{2}{\kappa^2}\sqrt{-g}R, \tag{2.31}$$

where $\kappa^2 = 32\pi G$, with G being the Newton's constant, and R the Ricci scalar. We take the background metric to be the Minkowski one, and linearize (2.31) taking the perturbation to be $\kappa h_{\mu\nu}$, so that $h_{\mu\nu}$ is the canonically normalized field. We get the following linearized Lagrangian:

$$\mathcal{L}_{\mathrm{EH}}^{(2)} = -\tfrac{1}{2}(\partial_\sigma h_{\mu\nu})^2 + \tfrac{1}{2}(\partial_\mu h)^2 + (\partial^\nu h_{\mu\nu})^2 + h\partial^\mu \partial^\nu h_{\mu\nu}, \tag{2.32}$$

where as usual we have introduced the notation $h := g^{\mu\nu}h_{\mu\nu}$, with $g_{\mu\nu}$ being the background (Minkowski) metric.

We note that (2.32) is much more involved than the gauge-theory Lagrangians we have encountered above. We also note that the trace part h of the metric perturbation behaves like a field with a wrong sign in front of its kinetic term. This is of course not a source of problems in GR, for this field turns out not to be dynamical. However, this does create problems for the path integral approach to the gravitational quantum theory, where one has to integrate over all metric fluctuations, including those described by h. This mode makes the Euclidean signature path integral not convergent. We will come back to this point below.

The above Lagrangian is invariant under the following transformations:

$$\delta_\xi h_{\mu\nu} = \partial_{(\mu}\xi_{\nu)}, \qquad (2.33)$$

which represent linearized diffeomorphisms.

The Hamiltonian description of the linearized GR is (briefly) as follows. Choosing the time+space split and decomposing the metric perturbation into its h_{00}, h_{0i}, h_{ij} components, we get (after quite some rearrangements and cancellations) the following involved structure:

$$\mathcal{L}_{\mathrm{EH}}^{(2)} = \tfrac{1}{2}(\dot{h}_{ij})^2 - \tfrac{1}{2}(\dot{h}_{kk})^2 - \tfrac{1}{2}(\partial_i h_{jk})^2 + \tfrac{1}{2}(\partial_i h_{kk})^2 + (\partial^i h_{ij})^2 + h_{kk}\partial^i\partial^j h_{ij}$$
$$+ h_{00}(\Delta h_{kk} - \partial^i\partial^j h_{ij}) + 2h_{0i}\partial_j p^{ij},$$

where we have introduced:

$$p^{ij} := \dot{h}^{ij} - \delta^{ij}\dot{h}_{kk}. \qquad (2.34)$$

It is clear that p^{ij} is the momentum canonically conjugate to h_{ij} and that the components h_{00}, h_{0i} have vanishing conjugate momenta and are Lagrange multipliers for constraints. One can then see that it is possible to set the trace part of p^{ij} by an action of a time-like diffeomorphism, and set the tracefree part $\tilde{h}_{ij}$ of h_{ij} to be transverse $\partial^i\tilde{h}_{ij} = 0$ by an action of a spatial diffeomorphism. Imposing these gauge-fixing conditions, we get the following simple Hamiltonian $\mathcal{H}_{\mathrm{EH}}^{(2)} = p^{ij}\dot{h}_{ij} - \mathcal{L}_{\mathrm{EH}}^{(2)}$:

$$\mathcal{H}_{\mathrm{EH}}^{(2)} = \tfrac{1}{2}((p^{ij})^2 + (\partial_i h_{jk})^2), \qquad (2.35)$$

where now both the metric perturbation and the conjugate momentum are symmetric traceless transverse tensors. We do not give all the details of this involved analysis; the main point of this discussion being to illustrate how much more complicated the case of metric-based GR is when compared with our treatment in the previous subsection. We have also verified that exactly the same Hamiltonian appears from the usual metric-based analysis and from our linearized Lagrangian (2.19). The Hamiltonian (2.35) resulting from the metric-based treatment, as well as the fact that the dynamical fields are symmetric traceless transverse tensors, is the reason why the graviton is referred to as a spin 2 particle. Having derived the same description of the graviton from the gauge theory Lagrangian (2.19), we have thus shown that the gauge-theoretic description of a spin 2 particle is possible.

Having reviewed the usual linearized GR situation, we are ready to discuss the question of the relation between our gauge-theoretic and the metric-based approaches. We have seen that once all the constraints are imposed and all the gauge freedom is fixed, one gets exactly the same Hamiltonian in both cases, the one describing two propagating polarizations of the graviton. Using the technical jargon, we can say that on-shell, the two descriptions of gravity are the same. The question is if they are the same off-shell, i.e. if there exists a field redefinition (possibly complicated) that maps (2.19) into (2.32). We will now give arguments that prove that there is no such a map, and so the two descriptions are different off-shell.

First, let us do a simple count of the number of components in the fields involved. The metric perturbation $h_{\mu\nu}$ contains 10 components, while the connection a_μ^i contains 12. However, in fact, as we have seen that the Lagrangian (2.19) depends only on eight of the 12 components of a_μ^i as it is invariant under the shifts (2.21) in the field space. It is certainly not possible to express 10 components of the metric perturbation in terms of eight components on which (2.19) depends.

Another reason why the Lagrangians (2.19) and (2.32) cannot be related is that they have very different properties in the field space. It is easiest to see this in the case when one passes to the Euclidean signature. Then there is no factors of i in the gauge-theoretic description, and the full connection a_μ^i is real. The self-dual two-forms in this case are also real. Then the gauge-theoretic Lagrangian is just the trace of the square of the real symmetric tracefree matrix $P^{ij|kl}\Sigma_k^{\mu\nu}\partial_\mu a_{\nu l}$. As such, it is explicitly non-negative. On the other hand, the Lagrangian (2.32), even when continued to the Euclidean signature, is still not positive-definite. Indeed, by choosing the conformal mode part h of the metric perturbation to be a field quickly changing in space–time one can make the Euclidean Lagrangian arbitrarily negative. This is well-known as the conformal mode problem in the Euclidean path integral. Thus, we are dealing with very different functionals in the two descriptions. In the gauge-theoretic description, the (Euclidean) Lagrangian is convex (apart from the flat directions corresponding to the gauge symmetries), while in the metric-based treatment, the functional is convex in some directions in the field space and concave in others (specifically the conformal mode direction). The critical point(s) of both functionals are the same, but it is clear that there cannot be any map from the field space of one description to that of the other that can relate two functionals with completely different properties. Thus, we conclude that the above gauge-theoretic description of linearized gravity is off-shell different from the description given by the Einstein–Hilbert action. This fact will have important implications when we discuss the quantum theory.

(f) Spinor description

The above description of the linearized theory, as well as the comparison with the metric-based approach, can be simplified quite considerably if one uses spinors. The main idea of the spinor description is that the space $S_+ \otimes S_-$, where $S_\pm$ are the spaces of the so-called unprimed and primed spinors (inequivalent two-dimensional representations of the Lorentz group), is isomorphic to the Minkowski space. This isomorphism is described by the object $\theta_\mu^{AA'}$, where we use the relativist notations for spinors and denote the spinor indices by upper case Latin

letters from the beginning of the alphabet. The spinors $\lambda^{A'}$ of opposite type to λ^A are called primed. This is in contrast to the particle physics notations where the spinor indices are Greek letters, and are referred to as undotted and dotted. The (hermitian in our conventions) object $\theta_\mu^{AA'}$ is called the soldering form. It encodes, in particular, the Minkowski metric via $g_{\mu\nu} = -\theta_\mu^{AA'}\theta_{\nu AA'}$, where the minus in this formula is convention (and signature)-dependent. Using the soldering forms one can convert any space–time index into a pair of spinor indices of opposite types. Thus, for example, the gauge field a_μ^a becomes $a_{AA'}^a$, where a is the Lie algebra index. For more information about spinors and their role in geometry, the reader is referred to Penrose & Rindler (1984).

We now describe the linearized Lagrangian (2.19) in spinor terms. As we shall see, most of the properties discussed above become manifest in this description. We already know that the space–time index of the connection a_μ^i becomes a pair AA'. We then recall that the self-dual two-forms $\Sigma_{\mu\nu}^i$ provide an isomorphism between the space of self-dual two-forms and the Lie algebra of SO(3), and take in the spinor language a very simple form. Indeed, a two-form $X_{\mu\nu}$ becomes in the spinor description an object $X_{AA'BB'}$. This is anti-symmetric under the exchange of pairs AA' and BB'. As such, it can be decomposed into either an object that is symmetric in AB and anti-symmetric in $A'B'$, or the object anti-symmetric in AB and symmetric in $A'B'$. Using the fact that the only AB anti-symmetric object is the metric in the space of spinors ϵ_{AB}, we have:

$$X_{AA'BB'} = \epsilon_{A'B'}X_{AB} + \epsilon_{AB}X_{A'B'}, \tag{2.36}$$

where X_{AB} and $X_{A'B'}$ are symmetric objects. Then, it can be shown that the $\epsilon_{A'B'}X_{AB}$ part of $X_{\mu\nu}$ is its self-dual part, and $\epsilon_{AB}X_{A'B'}$ is the anti-self-dual one. The objects $\Sigma_{\mu\nu}^i$ being self-dual are thus of the form $\epsilon_{A'B'}\Sigma_{AB}^i$. We can then use the objects Σ_{AB}^i to identify the Lie algebra index i with a symmetric pair AB of unprimed spinor indices. All in all, when all of the indices are replaced by the spinor ones, our linearized connection a_μ^i becomes the object $a^{BC}{}_{AA'}$.

Thus, we see that the linearized connection $a^{BC}{}_{AA'}$ is an object taking values in $S_+^2 \otimes S_+ \otimes S_-$, where the notation is that S_+^n is the space of symmetric rank n unprimed spinors. This space can be decomposed into its irreducible components:

$$S_+^2 \otimes S_+ \otimes S_- = S_+^3 \otimes S_- \oplus S_+ \otimes S_-, \tag{2.37}$$

where the completely symmetric $a^{(ABC)}{}_{A'}$ in its three unprimed indices part of $a^{BC}{}_{AA'}$ lies in $S_+^3 \otimes S_-$ and the trace part $a^{EA}{}_{EA'}$ lies in $S_+ \otimes S_-$. It is then easy to see that the part of the connection that can be shifted by the action of a diffeomorphism $\delta_\xi a_\mu^i = \xi^\alpha \Sigma_{\mu\alpha}^i$ is precisely the $S_+ \otimes S_-$ part. And indeed, the linearized Lagrangian (2.19) written in terms of spinors take the following extremely simple form:

$$\mathcal{L}_{GR}^{(2)} \sim (\partial_{A'}^{(A} a^{BCD)A'})^2, \tag{2.38}$$

where the numerical coefficient in front is convention-dependent and is not of importance for us here. As the symmetrization on all four unprimed indices is taken, the above Lagrangian is explicitly independent of the $S_+ \otimes S_-$ part of the connection. It is also easy to check that the Lagrangian (2.38) is invariant under

the gauge transformations $\delta_\phi a^{BC}{}_{AA'} = \partial_{AA'}\phi^{BC}$. Indeed, this follows from the fact that $\partial^{AA'}\partial^B_{A'} \sim \epsilon^{AB}\Box$ and thus contains only the AB anti-symmetric part, which gets killed by the symmetrization in (2.38).

It is worth emphasizing how much simpler is the description (2.38) when compared with all other known descriptions of linearized gravity. We shall give a comparison with the Plebanski–Ashtekar theory in the following subsection. Here, we would like to present a brief comparison with the spinor language metric case. In the spinor description, the metric perturbation $h_{\mu\nu}$ becomes $h_{AA'BB'}$, which is symmetric under the exchange of pairs AA' and BB'. As such, it decomposes into an object $h_{ABA'B'} \in S^2_+ \otimes S^2_-$, as well as the object $\epsilon_{AB}\epsilon_{A'B'}h$ proportional to the trace of $h_{\mu\nu}$. Thus, there is $9 + 1 = 10$ field components, and the Lagrangian (2.32) is a (complicated) function of all these fields. We note that even writing the Lagrangian (2.32) in the spinor form produces a rather cumbersome expression that is not particularly illuminating.

We have already remarked that the gauge-theory gravitational Lagrangian (2.19) is very similar to the Yang–Mills Lagrangian in the form (2.14). This similarity becomes even more striking in the spinor description. Thus, suppressing the Lie algebra index of the Yang–Mills field, we can write (2.14) as

$$\mathcal{L}^{(2)}_{YM} \sim (\partial^{(A}_{A'} a^{B)A'})^2. \tag{2.39}$$

Thus, the only difference between this Yang–Mills theory and (2.38) is that the former uses spin 1 fields in the representation $S_+ \otimes S_-$ (per Lie algebra generator), while the latter uses spin 2 fields in the representation $S^3_+ \otimes S_-$. Otherwise, both Lagrangians are constructed according to the same principle using the Dirac operator $\partial_{AA'}$. Indeed, in (2.39), the Dirac operator is used to map $\partial : S_+ \otimes S_- \to S^2_+$, and this symmetric rank two spinor is just squared to obtain the Lagrangian (and then the trace over the underlying Lie algebra is taken). In the case of (2.38), the principle is exactly the same except that the map is now $\partial : S^3_+ \otimes S_- \to S^4_+$, as is relevant for spin 2 particles. Below, we shall see that the similarity between this description of gravity and Yang–Mills theory extends beyond the linearized level, and that interaction vertices in our gauge theory description of gravity take very similar form to those familiar from the Yang–Mills.

Let us now see how the count of the number of degrees of freedom works in the metric and the gauge descriptions. The first difference is that the gauge theory Lagrangian is a function of only eight fields taking values in $S^3_+ \otimes S_-$, when compared with 10 fields from $S^2_+ \otimes S^2_- \oplus$ (trivial) in the metric case. In this sense, the gauge theory description is more economical than the metric-based one. The count of the propagating modes is then as follows. From the eight fields that the Lagrangian (2.38) depends on, three become Lagrange multipliers for three Gauss constraints, which gives the usual two propagating degrees of freedom. This is to be contrasted with the metric case count, where four out of 10 metric components are Lagrange multipliers for four constraints, which again give two propagating modes. The point that we would like to emphasize is that the gauge-theory description is less redundant—a smaller number of fields than in the metric case is sufficient to produce the same on-shell picture.

Second, the Lagrangian (2.38) is arguably a very simple and natural construct involving the basic field, in comparison with the metric Lagrangian that is not even easy to write down with all its four different terms. The final, crucial

difference comes from the fact that the Euclidean version of the Lagrangian (2.38) is positive-definite (being the trace of the square of a real symmetric tracefree matrix), while the Euclidean version of the linearized Einstein–Hilbert functional is not definite (because of the 'conformal' mode h). As we have already discussed above, this shows that the two descriptions can only be equivalent on-shell, while they are necessarily different off-shell.

Thus, all in all, the gauge-theory formulation (2.38) seems to be a much more economical and elegant description of gravity than the usual metric-based one. There is, however, a price to be paid for this simplicity. Indeed, the fields of the metric-based description take values in $S_+^2 \otimes S_-^2$ plus the trivial representation of the Lorentz group. In Lorentzian signature, the representations S_+ and S_- go into each other under the complex conjugation. Thus, the representation $S_+^2 \otimes S_-^2$ goes into itself under the complex conjugation, which explains why the metric-based description works with real fields. In contrast, in our approach, the basic field takes values in $S_+^3 \otimes S_-$, which under the operation of complex conjugation goes to $S_-^3 \otimes S_+$, a completely different representation. In other words, our description of gravity is necessarily chiral, with gravitons of one helicity being described in a different way from those of the other helicity. The chirality also implies that the gauge-theory description cannot be by real fields (apart from the cases of Euclidean $(+,+,+,+)$ and split $(-,-,+,+)$ signatures where the spinors are real objects). As the result, the issue of reality conditions becomes non-trivial.

(g) A comparison with the Plebanski–Ashtekar gauge-theoretic description

We close this section with a comparison between the above gauge theory description and that owing to Plebanski (1977) and (in the Hamiltonian formulation) Ashtekar (1987). A detailed description of the corresponding linearized theory is available in (e.g. Ashtekar (1991), ch. 11). Here, we will only give a very brief comparison, for more details, the reader is referred to the above references.

In Plebanski–Ashtekar description, one starts from a certain action functional that depends on an SO(3) connection A_μ^i, as well as an $\mathfrak{su}(2)$ Lie algebra-valued two-forms field $B_{\mu\nu}^i$, as well as a set of Lagrange multipliers. The precise form of this functional is not important for us here. What matters is the fact that it is not a functional of the connection only, as in our description, in that it depends on many other fields. In the Hamiltonian description many of these fields become non-dynamical. Thus, as usual, the temporal part of the connection becomes the Lagrange multiplier for the Gauss constraint. The temporal part B_{0j}^i of the B-field contains nine components, of which five are set to zero by constraints obtained by varying the action with respect to the Lagrange multipliers, and the remaining four are the Lagrange multipliers for the diffeomorphism constraints. The spatial part $b_{jk}^i \sim \epsilon_{jkl}\pi^{li}$ of the two-form field plays the role of the momentum conjugate to the spatial connection. The linearized theory constraints are then as follows:

$$\partial_i \pi^{ij} = 0 \ \text{(Gauss)},$$

$$\partial_{[i} a_{j]}^i = 0 \ \text{(Diffeo)} \tag{2.40}$$

and

$$\epsilon^{ijk}\partial_i a_{jk} = 0 \ \text{(Hamiltonian)},$$

where a_i^j is the spatial connection. The analysis of the propagating mode content of this system is quite non-trivial (it can be found in Ashtekar (1991), ch. 11), with the outcome being that the usual symmetric traceless transverse tensors are the propagating degrees of freedom.

Let us now compare the above with our gauge-theoretic description. The first difference is that our Lagrangian is a functional of the connection only, with no other auxiliary fields present. The second (crucial) difference is that the diffeomorphisms are realized in our description in a completely different way. Indeed, the spatial diffeomorphism as well as the Hamiltonian constraints (2.40) are non-trivial, and follow from varying the Plebanski action with respect to certain components of the temporal part of the B-field. In contrast, in our description, there are no fields in the action varying with respect to which one gets the diffeomorphism constraints. Instead, the action is simply independent of certain components of the connection. At the level of the Hamiltonian description, this is described by the statement that the following constraints hold:

$$\pi^{[ij]} = 0 \text{ (Diffeo)}, \quad \pi^{ii} = 0 \text{ (Hamiltonian)}. \tag{2.41}$$

As we have already said, these follow from impossibility to solve for certain components of the velocity in terms of the momenta, and not by explicitly varying the Lagrangian with respect to some Lagrange multipliers. What the constraints (2.41) generate are of course simple shifts of the connection variable in some directions. As a result, the analysis of the propagating mode content of (2.19) is much simpler than the corresponding analysis in the Ashtekar case.

All in all, our gauge-theory description is considerably different from the Plebanski–Ashtekar formulation, with both the linearized Lagrangian and Hamiltonian formulations being much more compact in our case. It is also worth emphasizing that the Plebanski Lagrangian, being essentially the Einstein–Hilbert Lagrangian with some extra fields added, is non-convex, while the (Euclidean) linearized Lagrangian (2.19) is explicitly convex. Thus, there are clear benefits from using the description (2.19) when compared with the Plebanski–Ashtekar formulation.

Finally, we note that our Lagrangian (2.19) can be obtained from that of Plebanski theory by the process of integrating out the two-form and the Lagrange multipliers fields, as was explained in Krasnov (2011*b*), and then considering the linearization of the resulting action, as will be explained below. So, Plebanski formulation of GR and the formulation that is the subject of this review are not unrelated. However, we will refrain from explaining this link in detail, as this will take us too far from the main subject of this paper, which is how to describe gravity by a theory with a connection as the only dynamical field.

3. Diffeomorphism-invariant gauge theories

In this section, we finally explain where the postulated above linearized gravity Lagrangian (2.19) comes from. We will also explain how both gravity and Yang–Mills theory arise from a single source—a general diffeomorphism-invariant gauge

theory. In this section, we switch gears completely, and follow a top-to-bottom approach, where we first present a general principle, and then specialize to the case of gravity plus Yang–Mills.

Some historical remarks are in order. While the gauge theories we encounter in this section are relatively new—in the form described here they have been introduced in works of the present author—they are intimately related to certain generally covariant theories of connection that has been discovered and studied in the early 1990s. In particular, the constructions of this section will make it clear that GR is not the only interacting theory of massless spin 2 particles. This fact has been known for quite some time in a different, but not unrelated formulation. The history of these developments is (briefly) as follows. It was realized in work (Capovilla *et al.* 1989) (see also Capovilla *et al.* 1991) that the zero cosmological constant GR can be reformulated as a theory of an SU(2) connection. The Lagrangian of this description, however, contains not just the connection field, but also an additional auxiliary field of density minus one. It was then quickly realized that in this formulation GR is not unique. Thus, a two-parameter family of 'neighbours' of GR was introduced and studied in Capovilla (1992), and then an infinite parameter of theories constructed along the same lines was proposed in Bengtsson (1991). All these theories are not ones of 'pure connection', as we have already said, they contain an additional auxiliary field. But they all have the key property that they describe just two propagating polarizations of the graviton. The same theories were rediscovered (in a different formulation) in Krasnov (2006) and this new version was related to earlier developments in Bengtsson (2007). The novelty of the approach followed here is that Lagrangians that are used are functions of only the connection field, with no auxiliary field being necessary. However, our theories can be though of as those studied in the 1990s with the auxiliary field integrated out.

(a) A class of gauge theories

Let us start by describing how gauge theory actions can be constructed in the absence of any space–time metric. This question is non-trivial, for up to now, we only know how to formulate a gauge theory (i.e. Yang–Mills theory) when a space–time metric is available (actually, since Yang–Mills is classically conformally invariant, only a conformal class of the metric is needed). It is not at all clear how the space–time indices of the curvature two-forms can be contracted in the absence of any metric.

To begin to answer this question, we note that the only object that can be used to contract the indices of copies of $F_{\mu\nu}^a$ and that is available without a metric is the tensor density $\tilde{\epsilon}^{\mu\nu\rho\sigma}$. This is a completely anti-symmetric object that in any coordinate system has components ± 1. We thus see that the following object with all space–time indices contracted is available:

$$\tilde{X}^{ab} := \tfrac{1}{4}\tilde{\epsilon}^{\mu\nu\rho\sigma} F_{\mu\nu}^a F_{\rho\sigma}^b, \tag{3.1}$$

where we have explicitly indicated that $\tilde{X}^{ab}$ is a density weight one (by putting a tilde over the symbol). The numerical prefactor of 1/4 is introduced for

convenience so that with our conventions $F^a \wedge F^b = \tilde{X}^{ab} d^4 x$. We now construct our Lagrangian by considering the most general gauge-invariant quantity that can be produced from (3.1).

Thus, the task is now to construct a scalar from (3.1). Moreover, this scalar must be a density one, so that it can be integrated over the space–time to produce an action. Thus, what we require is a function from the space $\mathfrak{g} \otimes_S \mathfrak{g}$, where $\mathfrak{g}$ is the Lie algebra of G and S denotes the symmetrization, to the real (or possibly complex) numbers, and this function must be homogeneous of degree 1:

$$f : \mathfrak{g} \otimes_S \mathfrak{g} \to \mathbb{R}(\mathbb{C}), \quad f(\alpha X) = \alpha f(X) \quad \forall X \in \mathfrak{g} \otimes_S \mathfrak{g}, \forall \alpha \in \mathbb{C}^*, \qquad (3.2)$$

where the last condition comes from our desire for $f(\tilde{X})$ to be of density weight 1. We also want our action to be gauge-invariant, and so as the gauge transformations act on the curvature two-forms by rotating the Lie algebra indices a (using mathematical terminology, the adjoint action), we also require

$$f(\mathrm{ad}_g X) = f(X) \quad \forall X \in \mathfrak{g} \otimes_S \mathfrak{g}, \forall g \in G, \qquad (3.3)$$

where $\mathrm{ad}_g X$ is the extension of the adjoint action of G in $\mathfrak{g}$ to $\mathfrak{g} \otimes_S \mathfrak{g}$.

One can then easily convince oneself that the class of such functions is not empty by giving examples. The first obvious example is obtained by taking $f(X) = \mathrm{Tr}(X)$. However, it is easy to see that the corresponding Lagrangian is a total derivative, and thus does not give rise to an interesting dynamical theory. A more non-trivial example is $f(X) = \mathrm{Tr}(X^2)/\mathrm{Tr}(X)$, which is homogeneous of degree 1 as required and gauge-invariant. In general, one can construct a very large set of functions with desired properties by taking all possible (independent) invariants constructed from $X \in \mathfrak{g} \otimes_S \mathfrak{g}$, and then taking an arbitrary function of homogeneity degree zero ratios of such invariants, and multiplying it by, say, $\mathrm{Tr}(X)$ to get the required total homogeneity degree 1. For our purposes, here it is enough to know that a very large class of functions with required properties exists.

Having discussed the properties of functions that can be used in constructing the action, we are ready to write down an action principle for a diffeomorphism-invariant gauge theory. Using the form notations, we have

$$S[A] = \mathrm{i} \int f(F \wedge F), \qquad (3.4)$$

where it is understood that the wedge product of two curvature two-forms is 4-form valued in $\mathfrak{g} \otimes_S \mathfrak{g}$, and the properties of function f make $f(F \wedge F)$ a well-defined 4-form that can be integrated over the space–time manifold to get the action. The imaginary unit i is introduced in front of the action for future convenience.

For future use, we note that the field equations following from (3.4) when this action is varied with respect to the connection are:

$$d_A \left(\frac{\partial f}{\partial \tilde{X}^{ab}} F^b \right) = 0. \qquad (3.5)$$

Here d_A is the covariant exterior derivative with respect to the connection A, and $\partial f / \partial \tilde{X}^{ab}$ is the matrix of partial derivatives of the function f with respect to the components of its matrix argument $\tilde{X}^{ab}$. We note that the matrix of

first partial derivatives is a function of $\tilde{X}$ that is homogeneous of degree 0. It is worth emphasizing that the field equations (3.5) are (nonlinear) partial differential equations that are second order in derivatives. The fact that only second-order differential equations appear is reassuring, for higher order field equations typically lead to instabilities.

It is instructive to compare (3.5) with the field equations following from the Yang–Mills action functional. These read $d_A{}^*F^a = 0$, where the dependence on the space–time metric enters via the Hodge star operation applied to F. Thus, the equations (3.5) are similar to those of the Yang–Mills theory, except that instead of applying to the curvature the Hodge star operator (which does not exist in our case), one contracts it with a matrix of first derivatives of the function f, which is itself depending on the curvature. Thus, equations (3.5) are well-defined in the absence of any space–time metric.

Importantly, we note that there are no dimensionful constants involved in the definition of the theory (3.4). Indeed, it is natural to take the dimensions of the connections to be those of $1/L$, L being length (or, using the standard terminology mass dimension 1). The quantities $\tilde{X}^{ab}$ are then of mass dimension 4, and owing to the homogeneity of f, so is the Lagrangian. Thus, there are only dimensionless constants involved in the construction of the Lagrangian of our theory, and these are hidden as the parameters of the function f.

Our final remark is about the generality of the above construction of a diffeomorphism-invariant gauge theory. We have considered actions that are built from the quantities (3.1). The natural question to ask is if there are any more involved gauge and diffeomorphism-invariant functionals of the connection that can be constructed. As we shall discuss in more detail below, this question is very important for understanding of the quantum theory based on (3.4), for it is related to the question of how the class of theories (3.4) behaves under the renormalization. To discuss this question, we note that a certain 'metric' tensor can be produced out of the curvature two-forms via:

$$\tilde{g}_{\mu\nu} = \tilde{\epsilon}^{\alpha\beta\gamma\delta} g_{ab} C^a_{cd} F^c_{\mu\alpha} F^d_{\nu\beta} F^b_{\gamma\delta}. \tag{3.6}$$

Here C^a_{cd} are the structure constants, and g_{ab} the Killing–Cartan metric on the Lie algebra $\mathfrak{g}$. The object $\tilde{g}_{\mu\nu}$ is a tensor density of weight 1, symmetric under the exchange of $\mu\nu$. It is a general group G analogue of the so-called Urbantke metric (Urbantke 1984). The choice of the proportionality coefficient in this formula is rather arbitrary, and one can, for example, always divide the object (3.6) by $\text{Tr}(\tilde{X})$ to get an object of density weight 0. So, in general (3.6) is only defined modulo rescalings. However, what is important for us here is that this object exists. In general, it is a non-degenerate tensor, which thus has an inverse. This inverse can then be used to contract indices of the operator of the covariant derivative $d_{A\,\mu}$ with that of the curvature $F^a_{\rho\sigma}$, producing a Lie algebra-valued object with one space–time index. This object can in turn be used to construct gauge and diffeomorphism-invariant objects of the type different from those considered before. Thus, the class of theories (3.4) describes only a subset of diffeomorphism-invariant gauge theories. This subset can be characterized by noting that for any actions other than (3.4), the resulting field equations are higher than second-order in derivatives. Thus, the class (3.4) consists of those diffeomorphism-invariant theories that lead to not higher than second-order field

equations. It is therefore certainly justified to restrict one's attention to such theories at low energies. A discussion of issues arising at high energies, when also the quantum effects become important, will be given below.

We would like to close this subsection with two references on somewhat analogous constructions of action functional that have appeared in the literature. Thus, Hitchin (ePrint) presents an action that is a functional of a p-form in n-dimensions. Such an action is constructed as a homogeneous of degree n/p function of a form. The result is a well-defined n-form that can be integrated to produce a 'volume' functional. The difference with our case is that we have constructed an action from Lie algebra-valued 2-forms (in four space–time dimensions), while in Hitchin (ePrint), non-trivial functionals arise for higher rank forms (3- and 4-forms) in higher dimensions. The forms in Hitchin (ePrint) are usual differential forms, and not forms with values in vector bundles as in our context.

Another relevant construction is Endlich *et al.* (2011), where an action principle for a perfect fluid is given in a form quite analogous to our (3.4). Here, one also works with symmetric matrices valued in some internal space, and constructs a gauge-invariant function of such matrices which is then integrated to produce an action. In addition, the requirement of invariance under the volume-preserving diffeomorphisms is imposed. The main difference with our context is that our basic field is a connection one-form, while in Endlich *et al.* (2011), it is a space–time scalar with values in some internal space. The perfect fluid construction of Endlich *et al.* (2011) suggests that the elementary excitations of our theory—gravitons—can be interpreted as analogues of phonons. Indeed, our gravitons arise as gapless modes once the original symmetries of the theory—diffeomorphisms and gauge rotations—are broken to a smaller (Poincare) subgroup by the background. We now turn to describing the linearization of (3.4) around an appropriately chosen background.

(b) Background

We start by noting that the construction (3.4) of a diffeomorphism-invariant gauge theory is empty in the simplest case $G = \mathrm{U}(1)$. Indeed, with the Lie algebra being one-dimensional, there is no other independent invariant of $\tilde{X}^{ab}$ except the trace (which in this case coincides with the quantity itself), and so the only possible theory in that case is the topological one (with the action being a total divergence). The next to most simple case is $G = \mathrm{SO}(3)$, which we shall now see describes gravity.

We will first explain how the Lagrangian (2.19) is obtained from (3.4) when the theory is linearized around a specific background. We will see that for this calculation, the specific form of the function f defining the theory is not important. Any choice of f satisfying a certain weak non-degeneracy condition will produce (2.19) via linearization. Then, in the next subsection, we describe how a specific choice of f gives GR.

We now need to discuss the background connection to be used for the linearization. We present a general description valid for any gauge group G, and then later specialize to the case $G = \mathrm{SO}(3)$. We first note that because ratios of invariants of $\tilde{X}^{ab}$ are used, we cannot take the zero (flat) connection with vanishing curvature. Indeed, the ratios of curvature invariants that are arguments

of the function f are then ill-defined. This is probably the most significant difference with the case of Yang–Mills theory, where the theory is completely well-defined around the zero connection. In our case, to obtain an expansion around the Minkowski background, there is no other choice but to resort to some sort of limiting procedure. A natural limiting procedure is to take a constant curvature connection and then send the radius of curvature to infinity, thus reproducing the Minkowski space–time. We shall see that this is a well-defined procedure with an unambiguous outcome. Alternatively, one can interpret this limiting procedure as that of finding the linearized description of the theory (3.4) for a constant curvature background, on scales much smaller than the curvature scale. We note that this is a physically realistic situation, because the metric of our Universe is not Minkowski, and, during the current late stages of the evolution of the Universe, is in fact quite close to the de Sitter space.

To describe the background connection, we shall require it to have a large degree of symmetry, so that it can legitimately be referred to as a 'vacuum' state of our theory. The symmetry requirements that we impose are those standard in cosmology: homogeneity and isotropy. Thus, we require that there exists a foliation of the space–time such that on each three-dimensional slice of the foliation, the connection is homogeneous, i.e. looks exactly the same at every point. This is formalized by introducing a time coordinate, which we for now shall refer to as η, so that $\eta = \text{const.}$ are the time-slices of homogeneity. The connection will then only be allowed to depend on η. We then require the connection to be spherically symmetric, i.e. such that the effect of spatial rotation around any point on $\eta = \text{const.}$ hypersurface can be removed by a gauge transformation. It is not hard to show that this fixes the connection to be of the form

$$A^a_{\text{backgr}} = \frac{a^a_i}{i} \, dx^i, \qquad (3.7)$$

where x^i are the usual $\mathbb{R}^3$ coordinates on $\eta = \text{const.}$ surfaces, a^a_i is a function of η only, and we have used the availability of time-dependent gauge transformations to eliminate the $d\eta$ components. The requirement of spherical symmetry further requires a^a_i to be an *embedding* of the Lie algebra $\mathfrak{so}(3) \sim \mathfrak{sl}(2)$ of the group of spatial rotations SO(3) into the Lie algebra $\mathfrak{g}$, i.e. a map that sends $\mathfrak{sl}(2)$ commutators into $\mathfrak{g}$ commutators. We refer the reader to the study by Krasnov (2011c) for a demonstration of this claim. In the case $G = \text{SO}(3)$, there is only one such (non-trivial) embedding, and so in this case a^i_j is proportional to δ^i_j, but in the case of a larger gauge group G, there are typically several inequivalent (i.e. non-conjugate in G) embeddings. Later, we shall see that it is very interesting to consider different embeddings in (3.7), as one gets very different particle spectra around non-equivalent 'vacua'.

Let us now rewrite the connection components a^a_i as

$$a^a_i = e^a_i a(\eta), \qquad (3.8)$$

where e^a_i is an embedding of $\mathfrak{sl}(2)$ into $\mathfrak{g}$ normalized so that $f^a{}_{bc} e^b_j e^c_k = \epsilon^i{}_{jk} e^a_i$, and $a(\eta)$ is an arbitrary function of time η. We shall now see that the only freedom in choosing the vacuum (3.7) is the choice of the embedding, and that the time

coordinate can always be conveniently chosen so that the dependence of a on time is fixed. Thus, let us compute the curvature of (3.7). We have

$$F^a = -a^2 e_i^a \left(i\frac{a'}{a^2} \, d\eta \wedge dx^i + \frac{1}{2} \epsilon^i{}_{jk} \, dx^j \wedge dx^k \right). \tag{3.9}$$

We now choose the time coordinate so that $a'/a^2 \, d\eta = dt$, which fixes the connection component a as a function of t:

$$a(t) = \frac{1}{t_0 - t}, \quad t < t_0, \tag{3.10}$$

where t_0 is an arbitrary integration constant.

Let us now inspect the background curvature (3.9) more closely. The expression (3.9) shows, in particular, that in the six-dimensional space of all two-forms, the background curvature components span a three-dimensional subspace

$$\Lambda^+ := \mathrm{Span}(F^a), \quad \dim(\Lambda^+) = 3. \tag{3.11}$$

We can now introduce a metric, or rather a conformal class of metrics, by requiring this subspace $\Lambda^+ \subset \Lambda^2$ to be that of self-dual two-forms with respect to the metric. It is known that this requirement fixes the metric modulo conformal rescalings. Let us stress that we are not forced to introduce a metric, as our theories are perfectly well-defined and can be studied in the absence of one. However, as all the physics as we know it happens in the metric background, it is very convenient to introduce a metric so that later, we have a chance to see the familiar physics arising. We are not yet claiming that the metric introduced via the above mathematical construction is a physical one, in the sense that it is the one that the matter fields feel. However, we shall later see that this is indeed the case.

Let us now discuss the ambiguity in the choice of the representative in the conformal class defined by declaring Λ^+ to be the self-dual two-forms. Given any such metric, let θ^I be the corresponding tetrads: $ds^2 = \theta^I \otimes \theta^J \eta_{IJ}$, where $I, J = 1, \ldots, 4$ and η_{IJ} is the Minkowski metric. We can then define the 'canonical' self-dual two-forms for this metric via

$$\Sigma^i := i\theta^0 \wedge \theta^i + \tfrac{1}{2}\epsilon^i{}_{jk}\theta^j \wedge \theta^k. \tag{3.12}$$

For metrics related by a conformal rescaling $g_{\mu\nu} \to \Omega^2 g_{\mu\nu}$, the corresponding self-dual two-forms Σ^i are related via $\Sigma^i_{\mu\nu} \to \Omega^2 \Sigma_{\mu\nu}$. Thus, any metric whose self-dual two-forms Σ are a multiple of

$$\Sigma^i_{\mathrm{Mink}} := i\,dt \wedge dx^i + \tfrac{1}{2}\epsilon^i{}_{jk}\,dx^j \wedge dx^k \tag{3.13}$$

is a possible metric in the conformal class fixed by (3.11). It is, however, very convenient to work with a metric such that

$$F^a = -M^2 e_i^a \Sigma^i, \tag{3.14}$$

where M^2 is a constant (i.e. time-independent) parameter. This is the de Sitter metric

$$ds^2 = c^2(t) \left(-dt^2 + \sum_i (dx^i)^2 \right), \tag{3.15}$$

of constant curvature M^2 or of the cosmological constant

$$M^2 = \frac{\Lambda}{3}. \tag{3.16}$$

Here

$$c(t) = \frac{1}{M(t_0 - t)}. \tag{3.17}$$

One could also consider anti-de Sitter space if one so wished by taking the parameter M to be purely imaginary. However, as the observational evidence points to the existence of a positive cosmological constant, we choose our background metric to be that of a de Sitter space.

The introduced parameter M of dimensions of mass is completely arbitrary, for one can always rescale the metric by a constant, while at the same time rescaling M^2 by the inverse constant without changing the background curvature F^a. We now remind the reader that our theories do not have any dimensionful constants. We will later see that once (3.4) is expanded around backgrounds (3.7), interpreted as equipped with a metric of constant curvature M^2, the dimensionful parameter M serves as a seed for all other dimensionful parameters in the theory, such as masses (for some massive fields that will be discussed below), and such as dimensionful coupling constants (such as e.g. the graviton interaction strength). Thus, the parameter M is in fact just a unit of mass for our theories. As such, it is inconsistent to ask about its value, for there is nothing else with which this value can be compared.

All our backgrounds (3.7) are constant curvature, in e.g. the sense of equation (3.14). Thus, the most symmetric solution of our theory (3.4) is naturally a Universe with a cosmological constant (in the sense of the de Sitter metric (3.15)). In this sense, a non-zero cosmological constant need not be explained in our formulation of gravity, it comes as a given. Moreover, as we have already stressed above, it is logically inconsistent to ask about its value, for being the only dimensionful parameter and that there is nothing with which this value can be compared. The cosmological constant is just a unit in which all other quantities get expressed.

The above discussion has very interesting implications. In our theories, the cosmological constant Λ, together with the usual $c, \hbar$, becomes the fundamental constant, such that any other constant of Nature can be expressed as a multiple of some combination of $\Lambda, c, \hbar$. Thus, for our class of theories, the cosmological constant Λ plays the same role as the Newton constant plays conventionally. In fact, we shall see that the Newton constant that measures the strength of interaction of gravitons does get expressed as a multiple (necessarily very large) of M^{-2}. When put in this way, the famous cosmological constant problem of why Λ is so small becomes in our case the question of why $M_p^2 \sim 1/G$ is so large (when compared with Λ). We will come back to this question below.

(c) Action evaluated on the background

Before we proceed with our analysis of the theories (3.4) linearized around the background (3.7) we make a small detour and compute the value that the action

takes when evaluated on (3.7). This small computation is related to the issue of other physical parameters being expressed in terms of the cosmological constant discussed above.

We have $F^a \wedge F^b = M^4 e_i^a e_j^b \Sigma^i \wedge \Sigma^j = 2\mathrm{i} M^4 e_i^a e_j^b \delta^{ij} \sqrt{-g} d^4 x$, where $\sqrt{-g} = c^4(t)$ is the square root of the determinant of the de Sitter metric. So, we have

$$S[A_{\text{backgr}}] = -2M^4 f(m) \int \sqrt{-g} \, d^4 x, \qquad (3.18)$$

where we have introduced a notation $m^{ab} := e_i^a e_j^b \delta^{ij}$ for the pullback of the metric on $\mathfrak{sl}(2)$ into $\mathfrak{g}$ by the embedding e_i^a. Comparing this with the value of the Einstein–Hilbert action evaluated on the constant curvature metric, we see that we must identify

$$2M^4 f(m) = \frac{\Lambda}{8\pi G}. \qquad (3.19)$$

In other words, the product $M^4 f(m)$ must be identified with the energy density of the cosmological constant. Now, using the above identification (3.16), we see that we must have

$$f(m) \sim M_p^2 / M^2. \qquad (3.20)$$

This is a very large number $M_p^2 / M^2 \sim 10^{120}$. Below, we shall see similar large numbers appearing if we want to get the usual graviton interaction strength from our approach. Thus, we learn that the functions f that are capable of reproducing the realistic physics are quite special in the sense that the above sort of scales appear. We will come to this point when we discuss what the cosmological constant problem becomes within our approach.

(d) General linearization

To see how (2.19) arises, let us first derive the general linearization formula, without specifying the $G = \mathrm{SO}(3)$ case. Thus, we will continue to use $a, b, \ldots$ for the Lie algebra indices. Now, let A_μ^a be (an arbitrary for now) background connection and δA_μ^a be a perturbation. The first variation of (3.4) reads:

$$\delta S = 2\mathrm{i} \int \frac{\partial f}{\partial \tilde{X}^{ab}} F^a \wedge D_A \delta A^b. \qquad (3.21)$$

The second variation reads:

$$\delta^2 S = \mathrm{i} \int 4 \frac{\partial^2 f}{\partial \tilde{X}^{ab} \partial \tilde{X}^{cd}} (F^a \wedge D_A \delta A^b)(F^c \wedge D_A \delta A^d)$$
$$+ 2 \frac{\partial f}{\partial \tilde{X}^{ab}} (D_A \delta A^a \wedge D_A \delta A^b + F^a C^{bcd} \delta A^c \wedge \delta A^d). \qquad (3.22)$$

Here C^{abc} are the structure constants. One must substitute the background curvature (3.14) (as well as the background connection) into this second-order action, and study the spectrum of the arising excitations.

(e) Case $G = \mathrm{SO}(3)$: linearized theory

We now note that in view of (2.7), the background value of the matrix $\tilde{X}^{ij}$ is proportional to δ^{ij}. Using this fact, we can say a great deal about the matrices of first and second partial derivatives of f present in (3.22). First, it is clear that the matrix of first derivatives can only be proportional to δ^{ij}. To see this, one can think about the function $f(\tilde{X})$ as being a power series expansion in traces of powers of $\tilde{X}$, as well as inverses of such traces. It is then clear that the value the matrix of first derivatives of such a function can take on δ^{ij} must itself be proportional to δ^{ij}. The proportionality coefficient is then a constant (because the background was chosen to be of constant curvature). We can then integrate by parts in the first term on the second line in (3.22) and, assuming that the connection perturbation vanishes outside of the region of interest, see that the result cancels with the second term. So, for the background considered, we can ignore the second line in (3.22).

Let us now consider the term involving the matrix of second derivatives of f. The form of this matrix can be determined from the fact that f is a homogeneous function of order one. This property can be summarized by writing

$$\frac{\partial f}{\partial \tilde{X}^{ij}} \tilde{X}^{ij} = f. \tag{3.23}$$

This identity can be differentiated with respect to $\tilde{X}^{kl}$ with the result being

$$\frac{\partial^2 f}{\partial \tilde{X}^{ij} \partial \tilde{X}^{kl}} \tilde{X}^{ij} = 0. \tag{3.24}$$

When the background value of $\tilde{X}^{ij}$ is proportional to δ^{ij}, this implies that the matrix of second derivatives of f gives zero when contracted with the Kronecker delta in any of its pairs of indices. This implies that the matrix of second derivatives is proportional to the projector on the symmetric tracefree tensors:

$$\left. \frac{\partial^2 f}{\partial \tilde{X}^{ij} \partial \tilde{X}^{kl}} \right|_{\tilde{X}=\delta} = -\frac{g}{2} P^{ij|kl}. \tag{3.25}$$

where g is some constant, capturing the content of the Hessian of the function f at $\tilde{X}^{ij} = \delta^{ij}$, and the numerical factor and the sign are introduced for convenience (we will later see that g is positive for GR). Note that $g = 0$ for the function $f(X) = \mathrm{Tr}(X)$ corresponding to the topological theory, but for a generic choice of f, it is non-zero. The Hessian at the point $\tilde{X}^{ij} = 2\mathrm{i}M^4\sqrt{-g}\delta^{ij}$ is easily determined using the homogeneity of f. All in all, collecting all the factors and dividing the second variation of the action by two to get the linearized action we get:

$$S^{(2)} = \frac{g}{2} \int \sqrt{-g} P^{ij|kl} (\Sigma_i^{\mu\nu} D_\mu \delta A_{\nu j})(\Sigma_k^{\rho\sigma} D_\rho \delta A_{\sigma l}). \tag{3.26}$$

Note that factors of M have cancelled. Now, to get the Lagrangian (2.19) we define a new, canonically normalized field $a_\mu^i = \mathrm{i}\sqrt{g}\delta A_\mu^i$. Note that there is a factor of the imaginary unit in this formula, as in e.g. (3.7). We now take the limit $M \to 0$, or, alternatively, consider the above linearized theory on scales much smaller than the scale of the curvature. In both cases, we can replace the covariant derivative

present in (3.26) by the ordinary derivative, and remove the volume density for the metric. We get:

$$S^{(2)} = -\frac{1}{2} \int P^{ij|kl} (\Sigma_i^{\mu\nu} \partial_\mu a_{\nu j})(\Sigma_k^{\rho\sigma} \partial_\rho a_{\sigma l}), \qquad (3.27)$$

which is the integral of the Lagrangian (2.19) studied above.

(f) Case $G = \mathrm{SO}(3)$: recovering full general relativity

We have just seen how for $G = \mathrm{SO}(3)$, the linearization of (3.4) around the constant curvature background (3.7) produces the gauge-theoretic gravity Lagrangian (2.19) studied in the previous section. This has happened independently of which function f was taken in the construction of the action. We shall now show that for a particular choice of f, the full on-shell content of (3.4) is the same as in GR. In other words, we will now show that for a specific f, there is a correspondence between solutions of GR and solutions of theory (3.4). As we have already discussed in the section on the linearized theory, we can at best hope to have an on-shell correspondence between the two descriptions, for the off-shell, even the linearized theories have completely different convexity properties.

Let us first state that the function f that corresponds to GR is given by Krasnov (2011b):

$$f_{\mathrm{GR}}(X) = \frac{1}{16\pi G\Lambda} (\mathrm{Tr}\sqrt{X})^2, \qquad (3.28)$$

where G is the Newton's and Λ is the cosmological constant. Note that, as discussed above, there are only dimensionless constants entering into the construction of f. We also note that the function f evaluated on the identity matrix $X^{ij} = \delta^{ij}$ is of the order M_p^2/M^2, which we knew we should have expected from general considerations leading to (3.20). We also note that the cosmological constant Λ enters f in the denominator, so the function f as well as the action blow up in the limit $\Lambda \to 0$. Thus, strictly speaking, only the non-zero Λ case GR is covered by our gauge-theoretic description. However, the flat case can be obtained by a limiting procedure of the sort we have used in the previous subsection.[1] We feel that it is not a drawback of our scheme that only the non-zero Λ case GR admits an honest action principle. Indeed, it is now commonly believed that there is a small cosmological constant in the Universe (or at least all the observational data are compatible with the dark energy being a cosmological constant). Thus, the action principle with f as above is sufficient to describe the physical Universe. Also, as we have said, if desired, the properties of the flat Universe can be recovered by a limiting procedure. In addition, as we shall indicate below, the present gauge-theoretic formulation may allow to explain the form of the 'physical' function f as an outcome of some sort of renormalization

---

[1]A 'pure connection' action functional that covers the $\Lambda = 0$ case is also possible, and is known as the Capovilla–Dell–Jacobson (CDJ) action (Capovilla *et al.* 1989). It is, however, not of the type (3.4) as it contains an additional auxiliary field. The CDJ action can be thought of as the δ-function imposing the condition that the trace of the square root of the matrix of wedge products of the curvature vanishes. The same condition follows from (3.28) in the limit $\Lambda \to 0$, and in this sense the $\Lambda = 0$ case is covered by our description via a limiting procedure.

group flow. Thus, the present approach may have something to say about the problem why Λ is so small—the famous cosmological constant problem. More remarks on this will be given below.

The notion of the square root used in (3.28) requires clarifications. As it can be shown, (see Krasnov 2011*b*) for more details, in the case of Euclidean signature GR, the matrices X that arise are squares of real matrices and therefore are real and not indefinite, i.e. all their eigenvalues are of the same sign (with some possibly being zero). For such matrices, there is a well-defined notion of the square root, and this is what is used in (3.28). In the case of the physical, Lorentzian signature GR, the situation is more complicated, because the matrices X are in general complex. However, even in this case, we can define the square root in the neighbourhood of the identity matrix by writing $\sqrt{X} = \sqrt{\delta + (X - \delta)}$ and expanding in powers of the small matrix $X - \delta$. This definition is sufficient for most practical applications, and certainly for the perturbative treatment of gravity. Thus, one can take the position that function (3.28) can be defined by a power series expansion for curvatures whose departure from the constant curvature case is not large. And in the large curvature case, where a perturbative definition would break down one cannot trust the theory of GR anyway, as higher order corrections to the GR Lagrangian (induced by e.g. quantum effects) must be taken into account. To summarize, the function (3.28) is well-defined in all situations where one is interested in classical GR predictions.

To show that the theory with (3.28) is on-shell equivalent to the usual GR with its Einstein–Hilbert action, let us first write down the field equations (3.5), specialized to the case of the defining function f given by (3.28). Let us first define, for any f

$$B^i := \frac{\partial f}{\partial \tilde{X}^{ij}} F^j. \tag{3.29}$$

The field equations (3.5) are then $d_A B^i = 0$. For the function (3.28), we have

$$\frac{\partial f}{\partial \tilde{X}^{ij}} = \frac{1}{16\pi G\Lambda} (\mathrm{Tr}\sqrt{\tilde{X}})(\tilde{X}^{-1/2})^{ij}, \tag{3.30}$$

where a negative power of the matrix $\tilde{X}^{ij}$ should also be understood in the sense of a power series expansion. It is then easy to check that precisely for the function (3.28), we have the following important identity satisfied by the two-forms B^i:

$$B^i \wedge B^j \sim \delta^{ij}. \tag{3.31}$$

One can reverse this argument and say that the function f in (3.28) is chosen precisely in such a way that the identity (3.31) holds. This identity is analogous to (2.7) already encountered above.

Overall, we get the following structure. First, any SO(3) connection A^i gives via (3.29) a set of two-forms B^i satisfying (3.31). It is then known that such a triple of two-forms defines a unique space–time metric $g_{\mu\nu}$. The metric $g_{\mu\nu}$ is determined by the requirement that the triple of two-forms B^i is self-dual with respect to $g_{\mu\nu}$ (this defines the metric modulo conformal transformations), and that the volume form of the metric coincides with a multiple of $\mathrm{Tr}(B \wedge B)$. For an explicit expression, one can use the Urbantke formula (3.6), specialized to the case $\mathfrak{g} = \mathfrak{su}(2)$, with the two-forms B^i substituted in place of F^i. Then for a connection satisfying its

field equation $d_A B^i = 0$, it can be shown that A^i coincides with the self-dual part of the Levi-Civita connection for the metric defined by B^i. Finally, the defining relation (3.29) can be read as saying that the curvature of the self-dual part of the Levi-Civita connection is self-dual as a two-forms, which is known to be equivalent to the Einstein condition, see proposition 2.2 in Atiyah *et al.* (1978), or Krasnov (2009) for an exposition oriented towards a physics audience. This establishes the on-shell equivalence of theory (3.4) with (3.28) with Einstein's GR. As we have already discussed, there is only the on-shell equivalence, i.e. one-to-one map between the spaces of solutions of field equations. Off-shell behaviour of the functionals (3.4) with (3.28) and the Einstein–Hilbert functionals are very different. See the study of Krasnov (2011*b*) for a more detailed discussion of the equivalence of the two theories.

Let us also discuss the relation in the opposite direction. Here one takes an Einstein metric and then computes the self-dual two-forms Σ^i given by (3.12), which satisfy $B^i \wedge B^j \sim \delta^{ij}$. One also computes the self-dual part A^i of the spin connection. It can then be seen that these objects satisfy $d_A B^i = 0$, as well as an equation $F^i = \Psi^{ij} B^j$, where Ψ^{ij} are arbitrary coefficients, with the only condition being that $\mathrm{Tr}(\Psi) \sim \Lambda$. This shows that the self-dual part of the Levi-Civita connection is our sought A^i, which satisfies all the relevant field equations. The relation between the two descriptions can be made much more explicit, but here we will refrain from going into more details. Some more details will be given in the next section where we discuss the quantum theory and, in particular, the graviton scattering.

(g) *Deformations of general relativity*

Let us now discuss what happens when the function f used in the construction of the Lagrangian is different from (3.28). As we have already mentioned, when $f(X) = \mathrm{Tr}(X)$, we get a theory without any propagating degrees of freedom. As we have also seen above via the linearization, for any generic choice of f, one gets a theory describing a massless spin 2 particle. Thus, choices of f different from (3.28) give rise to interacting theories of massless spin 2 particles that are different from GR. Thus, it appears that for a generic choice of f, the action (3.4) with $G = \mathrm{SO}(3)$ gives rise to a counterexample to the GR uniqueness theorems, which are often quoted as saying that GR is the only consistent theory of interacting massless spin 2 particles, see e.g. the very first paragraph of Dvali *et al.* (2011). The purpose of this subsection is to explain why the above construction of the gauge-theory gravitational theories in no way contradicts the known uniqueness theorems.

Let us start with a brief review of the GR uniqueness results. This review cannot be exhaustive, as unavoidably only the results known to the author will be mentioned. Historically, the first type of uniqueness results is based on an old idea that the nonlinearities of gravity can be reconstructed from the fact that the gravitons must couple (in a canonical way) to their own stress-energy tensor. This line of reasoning is best-known from the description given in Feynman *et al.* (1995). Another well-known uniqueness result is Hojman *et al.* (1976), which works at the Hamiltonian level and analyses the algebra of constraints of GR. Another approach to GR uniqueness is based on the graviton scattering amplitude kinematical constraints (Grisaru *et al.* 1975). Finally, the

last well-known approach that we mention is due to Wald (1986), and analyses the conditions for the infinitesimal gauge symmetries visible at the linearized level to integrate into a gauge symmetry of the nonlinear theory.

We now discuss in turn each of the methods for proving the GR uniqueness. Let us start with the derivations that use the fact that the graviton field $h_{\mu\nu}$ couples to its own stress-energy tensor. In more details, the idea here is that knowing that the spin 2 field $h_{\mu\nu}$ couples canonically to its own stress-energy tensor (which is computed from the linearized action), one can hope to reconstruct the full nonlinear action of gravity. This was achieved (to some extent) by Feynman *et al.* (1995), and a much cleaner argument using the first-order formalism was given by Deser (1970). A similar argument fixes the coupling of any type of matter to gravity. We note that even in this case, there are some controversies and ambiguities, as was recently emphasized by Padmanabhan (2008). Our next comment is that this type of derivations assume that the only way to describe a spin 2 particle is via a field $h_{\mu\nu}$ with a symmetric pair of space–time indices. As we have seen, this is not at all the case, and it is equally possible to describe spin 2 particles using a connection variable a^i_μ, where $i = 1, 2, 3$ is an SO(3) index. And the proofs based on the $h_{\mu\nu}$ graviton descriptions are simply inapplicable to the connection description.

One could then ask a question if a similar, not same, logic can be applied in the connection case. However, in this case, we are unable to say that the connection a^i_μ couples to its own stress-energy tensor (as could be derived from the linearized Lagrangian). Indeed, there is simply no notion of the stress-energy tensor in our context (as there is no metric, and so, no notion of the variation of the action with respect to the metric). The best one could hope for is that the connection a^i_μ couples to some current J^i_μ that can be obtained (in some canonical way) from the linearized Lagrangian. One could then iteratively determine the nonlinearities and completely fix the theory. However, we do not see any physical requirement for why the connection should couple to a certain canonical current, at least in the presently considered case of pure gravity with no matter fields. Indeed, in the case of the usual gravity, the requirement of coupling of $h_{\mu\nu}$ to its own stress-energy comes about because one knows how gravity couples to the other matter. The consistency of the field equations (at next to linear order) then require the gravitational stress-energy coupling as well. No similar argument can be made in our present context of pure gravity. The case of gravity plus matter will be commented on below when we discuss gravity plus Yang–Mills unification in our framework. Thus, the line of reasoning based on Feynman *et al.* (1995), Deser (1970) is not applicable directly, because a different field is used to describe the spin 2 particles (a^i_μ instead of $h_{\mu\nu}$). At the same time, an analogous argument would require some notion of the universal current J^i_μ for gravitons, and it is not clear what physical argument would guarantee its existence.

A more directly applicable line of thought (Wald 1986) makes no assumption that gravitons couple to their own stress-tensor. Instead, the idea is simply to analyse the possible nonlinear completions of some linear field equations. Then, if some gauge symmetry was present in the linear theory (and thus there was an associated Bianchi identity), then there should be a nonlinear version of the Bianchi identity (consistency condition). This consistency condition is then very restrictive and limits the types of possible gauge symmetries. In our case, this

type of analysis can be applied directly, taking into account the fact that the spin 2 particle is now described by a_μ^i instead of $h_{\mu\nu}$. One can then see quite easily that there are two types of Bianchi identities satisfied by the field equations (3.5). One of these is a differential identity

$$d_A d_A \left(\frac{\partial f}{\partial X^{ij}} F^j \right) = 0. \qquad (3.32)$$

Using the fact that the commutator of two covariant derivatives is the curvature, the above identity can be written as:

$$0 = \epsilon^{ikl} F^k \wedge \frac{\partial f}{\partial X^{lj}} F^j \sim \epsilon^{ikl} X^{kj} \frac{\partial f}{\partial X^{lj}}, \qquad (3.33)$$

where we have used the definition of the matrix X^{ij} as the wedge product of two curvatures. The identity is then a direct consequence of the gauge-invariance of the function f. Note that (3.32) is in fact three equations, and is a differential identity for the field equations. It is clear that this identity has its origin in the gauge-invariance of the action.

Another type of identity does not involve any derivatives of the field equations. It reads:

$$\iota_\xi F^i \wedge d_A \left(\frac{\partial f}{\partial X^{ij}} F^j \right) = 0, \qquad (3.34)$$

and, as we shall see, holds for any vector field ξ. Here ι_ξ is the interior multiplication with a vector field ξ (and, in (3.34), the curvature 2-form). To prove this identity, we use the 'usual' Bianchi identity $d_A F^i = 0$ to take the curvature 2-form F^j out of the brackets. The quantity $\iota_\xi F^{(i} \wedge F^{j)}$, where the brackets denote the symmetrization, is then proportional to $\iota_\xi (\text{vol}) X^{ij}$, where (vol) is the volume form used to define X^{ij} via $F^i \wedge F^j = (\text{vol}) X^{ij}$. So, the above identity reduces to

$$X^{ij} d_A \left(\frac{\partial f}{\partial X^{ij}} \right) = 0, \qquad (3.35)$$

which is a consequence of the homogeneity (of degree one) of the function f. It is clear that (3.34) is related to the property of the action being invariant under the diffeomorphisms. We also note that the above arguments are valid for any choice of the gauge group, provided ϵ^{ijk} in the proof of the first identity is replaced with the appropriate structure constant.

To summarize, we see that for any choice of f, the theory (3.4) satisfies two Bianchi-type identities, with one of them (3.32) being a differential identity involving the field equations, and the other (3.34) being simply a statement that the field equations are not linearly independent. The gauge symmetries of the theory are then the usual gauge rotations as well as diffeomorphisms. No additional requirements is placed on the allowed interactions. In particular, no restrictions are placed on the type of function f that can be used in construction of the Lagrangian, provided this function satisfies the requirements of gauge-invariance and homogeneity. No unique Lagrangian (such as that of GR) can follow from such arguments.

Another GR uniqueness claim (Hojman *et al.* 1976) is based on the canonical analysis. The main idea of the argument is that the phase space of any theory of gravity is that of pairs (spatial metric, conjugate momentum). One also knows that the spatial metric must appear in the result of the commutator of two

Hamiltonian constraints. This, together with some assumptions, allows one to fix the form of the Hamiltonian constraint and thus the gravitational dynamics. This argument is not applicable in our case because it can be shown by a direct analysis (see Krasnov 2008) that the canonical variable that arises in the case of theories under consideration is not the spatial metric. The algebra of diffeomorphisms is indeed the usual expected algebra, but in our case, the spatial metric appearing in the commutator of two Hamiltonian constraints turns out to be a complicated function of the canonical variables. It thus appears that the form of the Hamiltonian cannot be fixed via this type of argument in the case when the configurational canonical variable is not directly related to the (spatial) metric.

Finally, yet another argument for GR uniqueness Grisaru *et al.* (1975) is based on the analysis of the 4-graviton scattering amplitude and showing that it is essentially determined (up to a numerical factor) by the kinematical constraints only. However, this analysis assumes parity invariance of the amplitudes (to reduce the number of the independent amplitudes). This assumption is explicitly violated by a generic theory from our family, with only the theory with (3.28) being parity invariant. Thus, while it appears that there is indeed a unique parity invariant gravity (which is GR), there also exists a large family of interacting theories of massless spin 2 particles that are not invariant under parity.

Much more can be said about the classical behaviour of the deformations of GR, but we cannot go into such details because of the space restrictions. We just mention that the spherically symmetric (BH) solution can be obtained for an arbitrary f (see Krasnov & Shtanov (2008) and also Krasnov & Shtanov (2009) for a more informal discussion). We note that the study by Krasnov & Shtanov (2008) uses a certain equivalent (but not obviously so) formulation of the same class of theories (similar to Plebanski formulation of GR). A derivation of the spherically symmetric solution is also possible directly in the 'pure connection' framework presented here, but this will be spelled out elsewhere. We note in passing that a general theory from our family exhibits an interesting resolution of the singularity inside the black hole phenomenon, which is described in detail in the study of Krasnov & Shtanov (2008).

(h) Gravity-Yang–Mills unification

Our above discussion of the coupling of gravitons to themselves (and to other matter) has already made it clear that this coupling is not as straightforward as in the usual metric-based description. In particular, there seems to be no notion of some canonical stress-energy tensor through which all matter would couple (and gravitons would self-couple). It then appears that the only consistent way to couple matter to our gravity theory is by enlarging the gauge group in the action (3.4). As we shall soon see, an SO(3) subgroup of such a larger group will still describe gravity, while the part that commutes with this SO(3) will describe the Yang–Mills sector. Constructions of this sort where first spelled out (in the framework of Plebanski-like formulation) by Torres-Gomez & Krasnov (2010) and Torres-Gomez *et al.* (2010). For earlier work with similar aims (in the context of the Hamiltonian formulation), see Chakraborty & Peldan (1994). Here, we describe what appears to be a much simpler derivation.

Thus, let us take a general theory from the family (3.4), with some gauge group G. As we have already discussed above, a particularly nice set of backgrounds, or 'vacua' for our theory is provided by (3.7). Such a background connection selects some $\mathrm{SU}(2) \sim \mathrm{SO}(3)$ subgroup of G. Note that there are in general several inequivalent embeddings of $\mathrm{SU}(2)$ into a larger gauge group. The idea is to fix such an embedding and look at the spectrum of excitations that arise around this background. Let us first look at the linearized theory, i.e. an expansion of the action up to second order in the fields. It is then easy to see that the $\mathrm{SU}(2)$ part of the gauge field perturbation is still described by the linearized action (3.27), as no details of the derivation presented in the corresponding subsection get changed. So, we still get two propagating polarizations of the graviton in this sector.

Let us now discuss the sector of the theory described by δA^a with the index a being in the part of the Lie algebra of G that commutes with the fixed $\mathrm{SU}(2)$. We note that this part may be empty (this depends on the embedding chosen). Let us first consider the terms in the second line of (3.22). The first term becomes (after the limit $M \to 0$ is taken) a total derivative and can be dropped. In the second term, for the background chosen, the only non-zero values of F^a are in the $\mathrm{SU}(2)$ part of the Lie algebra. But by assumption, the commutator of the two connections in the part of the Lie algebra that commutes with $\mathrm{SU}(2)$ does not have any component in the $\mathrm{SU}(2)$ part. So, the second term in the second line of (3.22) is also zero. We are left only with the first line. Here, the matrix of the second derivatives of the function f must be evaluated on the background $X^{ab} \sim (\delta^{ij}, 0)$, where one has δ^{ij} in the $\mathrm{SU}(2)$ part of the matrix and zero everywhere else. Also, one is only interested in $\partial^2 f / \partial X^{ia} \partial X^{jb}$ part of the matrix of the second derivatives. One can easily convince oneself that this matrix can only be proportional to $\delta^{ij}\delta^{ab}$. Thus, we can write:

$$\frac{\partial^2 f}{\partial X^{ia} \partial X^{jb}} = \frac{\kappa}{2} \delta^{ij} \delta^{ab}, \tag{3.36}$$

where κ is some constant. Note that this holds only for the part of the Lie algebra that commutes with the $\mathrm{SU}(2)$, because for the part that does not commute, there are other possible invariants that can appear on the r.h.s. of (3.36). Collecting all the constants (and dividing the second variation by two), we get the following linearized Lagrangian:

$$S_{\mathrm{YM}}^{(2)} = -\frac{\kappa}{2} \int \delta^{ij} \delta^{ab} (\Sigma^{i\mu\nu} \partial_\mu \delta A_\nu^a)(\Sigma^{i\rho\sigma} \partial_\rho \delta A_\sigma^b), \tag{3.37}$$

where we already took the limit $M \to 0$. As in the $\mathrm{SU}(2)$ sector, the factors of M have cancelled out in this derivation, and the only effect of the limit is in replacing the covariant derivatives by the usual ones (and in removing the $\sqrt{-g}$ factor). We now rescale $a_\mu^a = \sqrt{\kappa} \delta A_\mu^a$ and get the usual linearized Yang–Mills action in the form (2.14) encountered in the beginning of this paper.

Thus, it is very easy to see how both gravity and Yang–Mills linearized theories arise from the 'mother' theory (3.4), once it is expanded around the background (3.7). We see that the first of the expectations spelled out in §1, namely that a gauge-theoretic reformulation of gravity may be of help for the problem of unification with the other forces, seems to be at least partially fulfilled.

(*i*) *Symmetry breaking*

As we described above, backgrounds (3.7) that can be used as 'vacua' for our theories are in one-to-one correspondence with embeddings of the Lie algebra $\mathfrak{s}l(2)$ into the full Lie algebra $\mathfrak{g}$. We have discussed the interpretation of the connection components charged under the part of the Lie algebra $\mathfrak{g}$ that commutes with the embedded $\mathfrak{s}l(2)$. Thus, we have seen that these describe massless gauge bosons. It can be shown (Krasnov 2011*c*) that the other connection components, i.e. in those directions in $\mathfrak{g}$ that do not commute with the embedded $\mathfrak{s}l(2)$ typically describe massive fields of non-zero spin.

A very interesting symmetry-breaking scenario then becomes available. One can change the 'vacuum' around which the theory is studied. In one vacuum, where the centralizer of the embedded $\mathfrak{s}l(2)$ is non-trivial, we have the corresponding massless gauge bosons. In a vacuum with trivial centralizer, all massless gauge bosons of the previous vacuum become massive particles. Thus, by choosing the background connection (and thus the embedding of $\mathfrak{s}l(2)$) appropriately, we can break the Yang–Mills gauge group as desired.

This symmetry-breaking mechanism is described in detail in Krasnov (2011*c*). Here, we restrict ourselves to just illustrating how this mechanism works on the example of $G = \mathrm{SL}(3)$. Note that we work with complex Lie groups, so we make no distinction between, e.g. $\mathrm{SU}(3)$ and $\mathrm{SL}(3)$. There are two inequivalent embeddings of $\mathrm{SL}(2)$ into $\mathrm{SL}(3)$. One is the obvious one that embeds 2×2 matrices in, e.g. the upper left corner of 3×3 matrices in $\mathrm{SL}(3)$. This embedding breaks the original symmetry down to $\mathrm{SL}(2) \times \mathrm{U}(1)$, where the latter is generated by the matrices $\mathrm{diag}(1, 1, -2)$. As we have seen in the previous subsection, components of the connection charged under this $\mathrm{U}(1)$ gauge group describe massless gauge bosons—i.e. a Maxwell electromagnetic potential. The other fields that arise in this background turn out to be massive spin 3/2 particles that are electrically charged. There are two such fields, oppositely charged. It is unusual to have fields of half-integer spin described by commuting connection components, and so one might worry about the Hamiltonian being not bounded from below. However, the Lagrangian for this fields is second order in derivatives (and is of the same type as all Σ-containing Lagrangians we have encountered above). It is not hard to choose the reality conditions so that the Hamiltonian is explicitly positive-definite, and so no problems of the type seemingly predicted by the spin-statistics theorem arise. For more discussion on these issues, the reader is directed to Krasnov (2011*c*).

Let us now consider the other inequivalent embedding. In this case, there is no non-trivial centralizer of the embedded $\mathrm{SL}(2)$ in $\mathrm{SL}(3)$, and all the gauge symmetry apart from the gravitational $\mathrm{SL}(2)$ is broken. The massive field content of this model is as follows. One finds a massive spin 3, as well as a massive spin 1 field. We can say that the massless gauge boson have absorbed one of the components of the massive fields of the previous background, and became massive. The other massive fields have rearranged themselves into a spin 3 field.

To summarize, we have all the ingredients of a symmetry-breaking mechanism, where by changing the background, we can break the gauge symmetry as desired, with massless gauge bosons of one background becoming massive in another. Many more examples of this mechanism are described in Krasnov (2011*c*), including those with a realistic Standard Model symmetry-breaking pattern. We refer the reader to this reference for details.

We finish this section by stating that, in spite of the fact that our description of gravity is so very far from the standard one that there is a legitimate worry, that one will never be able to couple any realistic matter to it, at least some types of matter can be coupled very naturally. Thus, we have seen that the gauge fields can be coupled just by enlarging the gauge group. Quite interestingly, massive particles can also be coupled in the same way. In particular, and this is far from obvious, spin 0 massive particle can also arise in this framework (see Krasnov (2011*c*) for an example). It remains to be seen whether the Standard Model fermions can be embedded into this framework. It currently seems that if this is to be possible, the anti-commuting fermionic variables must be described by the grading-odd elements of some super Lie algebra.

4. Quantum theory

In this section, we finally discuss whether our gauge-theoretic formulation of gravity sheds any new light on the problem of quantum gravity. We start with a description of how the perturbative quantum theory looks like in the new language.

(*a*) Interactions

Like in the metric-based GR, one can expand the full nonlinear action (3.4) around some background, and then attempt to compute quantum corrections to the theory by evaluating Feynman diagrams. If one wants the asymptotic states to exist (one usually does), one has to work around the Minkowski space–time background. One complication (that can be overcome successfully) in our case is that the Minkowski space–time connection is $A = 0$, and as such it is difficult to expand around (imagine having to expand the Einstein–Hilbert action around the zero metric). So, the natural strategy appears to take the constant curvature background, and expand the action around it, taking the flat limit afterwards. We have already seen how the linearized gravity action (3.27) appears this way. One can continue this process, and compute the interaction vertices. It turns out, however, that one should be very careful about taking the limit $M \to 0$. Indeed, one finds, e.g. that the cubic interaction vertex for our gravitons is of the schematic form:

$$\mathcal{L}^{(3)} \sim \frac{1}{MM_p}(\partial a)^3 + \frac{M}{M_p}(\partial a)a^2. \tag{4.1}$$

Here, we have considered the case of the theory with f given by (3.28) that corresponds to GR. Then M_p is the Planck mass $M_p^2 \sim 1/G$. For a general f, the cubic interaction is schematically the same, except that one has $1/M^2$ in front of the first term, multiplied by some dimensionless combination of the parameters of f (see below). We remind the reader that the parameter M is the length scale introduced by the background (3.14), which is related to the cosmological constant as in (3.16). For details of this derivation, the reader can consult (Krasnov 2011*a*) (or perform a straightforward analysis him/herself).

There are several points to be noted about (4.1). First, there are non-renormalizable (by power-counting) interactions, such as the first term in (4.1). Second, the interactions seem to blow up in the limit $M \to 0$. Finally, unlike in

the metric-based GR, we get more than two derivatives in the interaction vertices. This can be easily seen to be a general feature, and the order n vertex will have as many as n derivatives in the form $(\partial a)^n$. We can write:

$$\mathcal{L}^{(n)} \sim \frac{1}{(MM_p)^{n-2}}(\partial a)^n + \cdots, \tag{4.2}$$

where the dots denote terms that have a lower number of derivatives. We also note that the number of derivatives in each successive term changes by two (as in (4.1)), so the next term in (4.2) is proportional to $(\partial a)^{n-2}a^2$. Thus, the perturbation theory is seemingly very different from the standard one.

Both of these puzzles (blowing up interactions and the higher number of derivatives) can be understood by noting that there is a relation between the connections and the metric perturbations, which is valid on-shell only, and which reads, schematically,

$$h = \frac{1}{M}\partial a. \tag{4.3}$$

There is yet another relation which is also valid on-shell, and which reads:

$$a = \frac{1}{M}\partial h. \tag{4.4}$$

Both of these can be true at the same time only on-shell $\partial^2 = M^2$. These relations can be used, in particular, to determine the helicity spinors in the connection description from those in the metric description.[2]

We see that it is tricky to take the limit $M \to 0$. This certainly cannot be done at the very start of the computation, because the interaction vertices seem to diverge in this limit, and so do the helicity states determined from e.g. (4.4). Thus, what one has to do at intermediate stages of the calculation is to work on the mass-shell $\partial^2 = M^2$, and then only take the limit $M \to 0$ after the physical quantities (such as the graviton-scattering amplitudes) are computed. This procedure can be shown to lead to the well-known graviton scattering amplitude results.

Let us describe in a bit more detail how e.g. the 4-graviton scattering amplitude can be reproduced. As in the metric-based GR, one can show that the 4-valent interaction vertex cannot contribute to the answer (in the helicity basis). Thus, one has to analyse only the contribution from diagrams involving two 3-valent vertices. The 3-valent interaction vertex is of the schematic form (4.1). However, let us make the structure of the expression in (4.1) more explicit. The spinor formalism is again most suited for this purpose. So, we convert all space–time as well as internal indices into spinor ones. Recall that the connection becomes an object $a^{ABCA'}$ with three unprimed and one primed spinor index. The relation (4.3) then reads:

$$h_{ABA'B'} = \frac{1}{M}(\partial a)_{ABA'B'}, \tag{4.5}$$

---

[2]The helicity spinors can of course be built independently from any metric description, but the two ways of determining the helicity states agree.

where we do not worry about constants (which are in any case convention-dependent). We used the following notation for the spinor contractions:

$$(\eth a)_{ABA'B'} := \eth^E_{(A'} a_{B')EAB}, \quad (\eth a)_{ABCD} := \eth_{E'(A} a^{E'}_{BCD}). \tag{4.6}$$

The second of these is to be used in the cubic interaction vertex below. In both expressions, here we assume the connection on the r.h.s. to take values in $S^3_+ \otimes S_-$, i.e. be totally symmetric in its three unprimed indices. As we have discussed in the section on linearized theory, this is the way to project out the components of the connection that are pure (diffeomorphism) gauge. The object on the left hand-side in (4.5) is then AB and $A'B'$ symmetric, and is the spinor representation of the tracefree part of the graviton field $h_{\mu\nu}$.

Let us also give the spinor expression for the 3-valent interaction vertex that one gets for the case of f corresponding to GR. Using the notations (4.6), we have:

$$\mathcal{L}^{(3)} \sim \frac{1}{MM_p}(\eth a)^{ABCD}(\eth a)_{ABE'F'}(\eth a)_{CD}{}^{E'F'} + \ldots, \tag{4.7}$$

where the dots denote terms that go to zero in the limit $M \to 0$. Using this vertex (and appropriate spinor helicity states for the external gravitons), one can reproduce many of the GR graviton-scattering amplitudes. In particular, the usual 4-graviton scattering amplitude can be obtained, as well as the general MHV amplitude (it can be shown that only this vertex contributes to the MHV amplitudes). The latter is obtained using essentially the same Berends-Giele-type recursion relations (Berends & Giele (1988)) that work in the case of MHV amplitudes of gluon scattering in Yang–Mills theory. Details of these computations will appear in a future publication.

Before we compare the above vertex with that in the case of metric-based GR, let us note that it is analogous to the cubic vertex of Yang–Mills theory (with the Lie algebra index suppressed):

$$\mathcal{L}^{(3)}_{\mathrm{YM}} \sim g_{\mathrm{YM}}(\eth a)^{AB} a_{AE'} a_B{}^{E'}, \tag{4.8}$$

where $(\eth a)_{AB} := \eth_{E'(A} a_{B)}{}^{E'}$ is essentially the self-dual part of the curvature, and g_{YM} is the Yang–Mills coupling constant. The main difference between our gravity vertex (4.7) and the above Yang–Mills one is that the latter is renormalizable while the former is not. This is related to two extra derivatives being present in (4.7) when compared with (4.8). The point that we would like to emphasize however is that the gauge-theory formulation graviton interaction vertex is as simple as that of Yang–Mills theory, which gives hope that computational complexity can be greatly reduced using this formalism.

Let us now quickly relate the vertex (4.7) to the cubic vertex in the metric-based formalism. We will only show the schematic form, because even the translation of (4.7) to the metric variables using the on-shell relations (4.3) and (4.4) is rather involved. We get:

$$\mathcal{L}^{(3)} \sim \frac{1}{M_p}(\partial^2 h)hh, \tag{4.9}$$

which is the usual 3-valent vertex of the metric-based GR. In short, this is the explanation how such a seemingly different theory as the one with interactions (4.7) can reproduce the usual graviton-scattering amplitudes.

Further, it is interesting to note that in the connection description, the parity invariance of GR does not become manifest. The parity invariance can be shown to hold (at least order by order in perturbation theory) for the theory (3.28). But it can also be seen to be explicitly violated by any other general member of the family (3.4). Thus, e.g. for the 4-graviton scattering in a generic theory, one finds that the all minus and all minus one plus amplitudes are zero, while the all plus and all plus one minus amplitudes are not. In the usual case, only the MHV amplitude $(+ + --)$ is non-zero for the 4-graviton scattering. So, the general member of the deformations of GR family can be seen to violate parity quite explicitly.

In preparation for a discussion of the quantum theory, let us briefly describe the structure of the interactions for a general f. In the case of f corresponding to GR, there are certain delicate cancellations which guarantee that there are no $\mathrm{Tr}((\partial a)_{ABCD})^3$ term in (4.7). For a general function f, there is no such cancellation, and the structure of the 3-vertex is

$$
\begin{aligned}
\mathcal{L}^{(3)} \sim \; & \frac{c_1}{M^2}(\partial a)_{AB}{}^{CD}(\partial a)_{CD}{}^{EF}(\partial a)_{EF}{}^{AB} \\
& + \frac{c_2}{M^2}(\partial a)^{ABCD}(\partial a)_{ABE'F'}(\partial a)_{CD}{}^{E'F'} + \cdots .
\end{aligned}
\tag{4.10}
$$

Note that the quantity M_{p} no longer appears, as there is no notion of the Planck mass for a general theory from the family (3.4). Indeed, recall that there are only dimensionless parameters in any of our theories (3.4). Thus, the quantities c_1, c_2 above are dimensionless parameters related to the derivatives of the function f evaluated at the identity matrix, and M, as before, is the length scale introduced by the background. For the f corresponding to GR, we have $c_1 = 0$, $c_2 \sim M/M_{\mathrm{p}}$, but for a general theory these are arbitrary dimensionless parameters. The relations (4.5) and (4.4) are still valid, and so the leading term of the 3-vertex can be rewritten in the metric terms as:

$$
\mathcal{L}^{(3)} \sim \frac{c_1}{M^5}(\partial^2 h)^3 + \frac{c_2}{M}(\partial^2 h)hh + \cdots .
\tag{4.11}
$$

The point of this exercise is to exhibit explicitly that the interaction vertices of a general theory from the family (3.4) contain, when rewritten in the metric language, terms with more than two derivatives. In general, the n-valent vertex will start from a term $(\partial^2 h)^n \sim (\mathrm{Riemann})^n$ when rewritten in the metric language. This fact will be important below when we discuss how the theories (3.4) behave under the renormalization.

(b) Renormalization

So far, we have only talked about the tree-level graviton-scattering amplitudes. These can be expected to agree with the results obtained in the usual metric-based approach (for theory (3.28)). However, as we have already discussed, the theory (3.28) is only on-shell equivalent to that described by the Einstein–Hilbert Lagrangian. Off-shell, the two actions have very different properties. Thus, one

can expect that the quantum theory based on (3.28) is different from that arising in the metric approach. More specifically, one should not expect the loop-corrected graviton-scattering amplitudes to be the same in the two approaches.

With this general remarks being made, let us discuss the issue of non-renormalizability of GR. In the usual metric formalism, it manifests itself on-shell starting at the two-loop order (Goroff & Sagnotti 1986). At this order of perturbation theory, one finds it necessary to add a new (Riemann)^3 term to the Einstein–Hilbert Lagrangian to cancel the arising divergence. This is expected to continue at higher loop orders (even though nobody has calculated the 3-loop divergences in pure gravity explicitly). It is thus believed that an infinite number of higher derivative terms has to be added to the Einstein–Hilbert Lagrangian in the process of its renormalization. This introduces an infinite number of new coupling constants, all becoming equally important around the Planck scale, and implying loss of predictive power of the theory at Planckian energies.

Let us discuss how the non-renormalizability of gravity can manifest itself in the gauge-theoretic description. We have seen that (some of) the coupling constants of our theory, once the Lagrangian is expanded around a constant curvature background, are of negative mass dimension, as e.g. in the first term in (4.1). It is thus to be expected that higher derivative counter-terms will get produced in the process of renormalization of the loop corrections. However, unlike the case with the metric-based GR, where the perturbative expansion of the Lagrangian only contains two-derivative vertices, the vertices in our description already contain arbitrarily high powers of the derivative (see (4.2)). This fact will be very important below.

Before we discuss implications of the higher derivative nature of our perturbative expansion, we emphasize that the appearance of all powers of the derivative does not mean that the theory is non-local. Indeed, as we have discussed in the previous section, the Lagrangian of the theory is nonlinear, but leads to not higher than second order in derivatives field equations. So, there is no non-locality in the theory, but, rather just nonlinearity that produces arbitrarily high powers of (∂a) when the Lagrangian is expanded.

The main point is then that, as arbitrarily high powers of the derivatives are already present in the expanded Lagrangian, the higher derivative terms that get produced in the process of the renormalization may already be contained in the original Lagrangian. Indeed, we have already seen how the 3-vertex (4.11) of the general theory, when rewritten in the metric language starts from the term $(\partial^2 h)^3$, which is nothing else but (Riemann)^3. This is exactly the counter-term that is known (Goroff & Sagnotti (1986)) to be needed in the usual metric-based quantum gravity at two loops. Thus, at least to the lowest order, the idea that the terms that have to be added to our Lagrangian in the process of renormalization are already contained in it seems to hold. It is far from being trivial that this continues to happen to all orders, but this can in principle happen, as it contradicts nothing we know about the behaviour of non-renormalizable theories. The important point here is not to see that even the higher derivative terms appear, for we have seen that the general n-valent vertex will start from the term (Riemann)^n. What would be non-trivial is that it is precisely the structure that can be encoded into a general Lagrangian of the form (3.4) that gets preserved by the process of the renormalization. For a more explicit calculation showing that

the (Riemann)[3] term that is known to arise in usual gravity at two-loops does appear in the perturbative expansion of a general Lagrangian from the family (3.4), we refer the reader to Krasnov (2010).

To summarize, thanks to the peculiar higher derivative structure of the perturbatively expanded Lagrangian of our theory, it may be that the class of theories (3.4) is closed under the renormalization. As we have already said, this is far from being trivial, for we know that the class (3.4) is not the most general class of diffeomorphism-invariant gauge theories. Indeed, as we have discussed in the previous section, the theories that lead to higher than second-order field equations are not included into (3.4). We have then emphasized that 'not higher than second-order field equations' is not equivalent to 'not higher than two derivatives in the interaction vertices', for the nonlinearities of a second-derivative field equations Lagrangian can (and do) manifest themselves as higher derivative interactions. The conjecture is then that in the process of renormalization of (3.4), only similar nonlinearities appear, but no terms that change the order of the field equations.

This conjecture was first stated in Krasnov (2006) in a different (Plebanski-like) formulation of this class of these theories. It is certainly non-trivial to prove, and the ongoing attempt is to compute the one-loop correction with the general Lagrangian (3.4) using the background field method. If it is found that at least at one-loop, one does not produce anything that is not already contained in (3.4), this would give a partial support to the conjecture. In the next subsection, we shall discuss a possible scenario for the ultraviolet (UV) behaviour of gravity in case the conjecture discussed in this subsection is correct.

(c) Conjectural RG flow

In this subsection, we shall assume that the class of theories (3.4) is closed under the renormalization and discuss what this implies for the problem of finding the UV completion of gravity. The closedness under the renormalization can be expressed as a property that the beta-function governing the renormalization group flow of the Lagrangian (3.4) is of the same form as the Lagrangian itself:

$$\frac{\partial f(F \wedge F)}{\partial \log \mu} = \beta_f(F \wedge F). \tag{4.12}$$

Here $\beta_f(F \wedge F)$ is a function with the same properties as f, i.e. a gauge-invariant, and homogeneous function of degree 1 of a symemtric 3×3 matrix, which can thus be applied to a matrix-valued 4-form. The subscript f indicates that this function depends on the function f itself (i.e. the beta-function is a function of the coupling constants encoded in f).

The UV behaviour of the theories (3.4) is then controlled by the UV fixed point(s) of the RG flow (4.12). One is then in the domain of applicability of the asymptotic safety ideas (see, Weinberg (2009) for a recent description). The idea of the asymptotic safety is that there is a UV-fixed point of the relevant RG flow (such as, e.g. (4.12)), with only a finite number of attractive directions. In other words, the idea is that most of the RG trajectories going to the UV approach the relevant fixed point, but then miss it and escape to infinity in the space of theories (coupling constants). And only a finite-dimensional subset of the trajectories actually end at the fixed point. The fixed point in question then
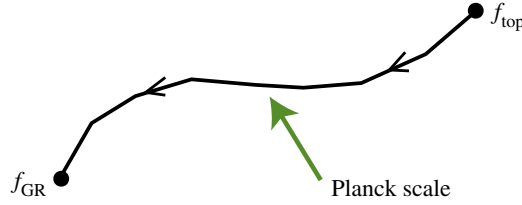
Figure 1. The conjectural RG flow. (Online version in colour.)

provides the UV completion of the theory, and one gets predictive power from the requirement that the physical theory is on one of the trajectories that leads to this fixed point. Then, as the dimension of the surface of such trajectories is finite, one needs only a finite number of measurements to fix the theory.

If the RG flow for the class (3.4) is of the form (4.12), the ideas of asymptotic safety can be applied directly. The flow is then essentially just a flow in the space of functions of two-variables $f(X) = f(\lambda_1, \lambda_2, \lambda_3) = \lambda_1 f(1, \lambda_2/\lambda_1, \lambda_3/\lambda_1)$, where we have used the fact that a gauge-invariant function of a symmetric 3×3 matrix X^{ij} is a function of its eigenvalues $\lambda_1, \lambda_2, \lambda_3$, where $X^{ij} = O \operatorname{diag}(\lambda_1, \lambda_2, \lambda_3) O^{\mathrm{T}}$ and $O \in \mathrm{SO}(3)$ is the diagonalizing orthogonal transformation. Such a flow can be studied very explicitly, and the question of whether there are any UV-fixed points with a finite number of attractive directions becomes answerable.

We then note that the ideas of asymptotic safety have a potential of explaining why it is the function (3.28) that appears as the low-energy Lagrangian. Indeed, one could envisage that being on a trajectory that emanates from the UV-fixed point necessarily implies that one flows to (3.28) at low energies. This way one could even explain the fact that the dimensionless constant in front of the GR Lagrangian (3.28), namely $1/(16\pi G\Lambda) \sim 10^{120}$, is so large. Indeed, it is not impossible that the large numbers of this sort are produced by the RG flows. Thus, it may be that our scenario has something to say about the very difficult cosmological constant problem, which is to explain why such two different scales seem to coexist in Nature.

Returning to the problem of the UV completion of gravity, we note that there is one point in the space of theories (3.4) that should certainly be a fixed point of the flow (4.12). This is the topological point $f_{\mathrm{top}}(X) = \operatorname{Tr}(X)$. Indeed, while this point is not strictly speaking among the class of theories that we consider (dynamically non-trivial theories with two propagating degrees of freedom), it should certainly annihilate the beta-function

$$\beta_{f_{\mathrm{top}}}(F \wedge F) = 0. \tag{4.13}$$

The reason for this expectation is that a topological theory without any dynamics can have no non-trivial running, and this should imply (4.13). It can be conjectured that the fixed point (4.13) is the sought UV-fixed point. If this is the case, then the dynamics in the UV region is controlled by this fixed point, with the 'safe' theories approaching this fixed closer and closer as one increases the energy. The arising conjectural picture of the UV behaviour of gravity is then depicted in figure 1.

5. Conclusions

We have come a long way: from a simple Lagrangian (2.19) (see also (2.38)) that describes free gravitons in Minkowski space–time, to a possible scenario of the UV completion of quantum gravity. We have also seen how both spin 2 particles (gravitons) and spin 1 particles ('gluons') are simultaneously described by the same theory (3.4), once it is expanded around an appropriate background. Here we would like to briefly reiterate the main ideas of the approach of this paper, and then list questions that remain open.

One of the main ideas was to use a gauge field rather than the metric as the main dynamical variable for gravity (in particular GR). We have seen how this is possible at both the linearized level (2.19), and, more non-trivially, at the full nonlinear level via theories (3.4) with (3.28). We have seen that gravity formulated in this language gets naturally embedded into a much larger class of theories (3.4), which we refer to as diffeomorphism-invariant gauge theories. The gravitational sector of such a general theory is then described by an appropriate SU(2) subgroup, while the other parts of the gauge group describe matter, in particular, the part of the gauge group that commutes with the gravitational SU(2) describing the Yang–Mills field.

Our construction of the class of theories (3.4) seems to be a very natural one, as it only uses the principles of diffeomorphism and gauge-invariance (which are both known to be of fundamental importance) as the input, as well as the desire to have the field equations of not higher than second order in the derivatives. We note that, importantly, there are no dimensionful parameters in our class of theories, with the dimensionful constants arising only when the theory is perturbatively expanded. This feature of theories (3.4) of not having any dimensionful couplings is, in our opinion, a very desirable one for any would be fundamental theory.

Importantly, our theories are nonlinear, while still leading to just second-order field equations. Nonlinearity is of course unavoidable in the real world, for without it there would be no interactions between particles. The most interesting theories that play a role in the description of Nature—Yang–Mills and GR—are both nonlinear. However, the nonlinearity of (3.4) is much stronger. In the case of Yang–Mills theory, the Lagrangian is polynomial in the fields (there are only up to quartic interaction vertices). In GR, non-linearity is worse, because once expanded (around a background) the action of GR contains vertices of arbitrarily high valency. Still, the order of the derivative that appears in GR vertices is always two (or smaller if one also takes into account the vertices coming from the cosmological term). The nonlinearity of our theories is similar to that in GR in that interaction vertices of arbitrary valency appear. However, it can be said to be stronger in the sense that up to the nth power of the derivative can appear at the n-valent vertex. On-shell, there is no contradiction between the two approaches. Indeed, as we have discussed, on-shell, the connection variable is related to the derivative of the metric (see (4.4)). This explains how what looks like an expansion in powers of ∂a can be rewritten as an expansion in powers of the metric perturbation. However, as we have discussed in detail in the first section of this paper, there can be no off-shell relation between the gauge-theoretic and the metric descriptions. Therefore, off-shell, and in particular, in quantum mechanical computations, one has to take the peculiar form of the perturbative

expansion (4.2) of the gauge-theoretic gravity Lagrangian seriously. It is then that the most interesting prospect, which may lead to an eventual UV completion of these theories, arises.

Both GR and our gauge theories (3.4) are non-renormalizable. But in GR, the non-renormalizability manifests itself in a very drastic way with new terms being added to the Lagrangian at every loop order to cancel the arising divergences. These terms are of higher order in the derivatives, and for this reason they could not have been in the original second derivative Einstein–Hilbert Lagrangian. In the case of theories (3.4), we can also expect UV divergences that will have to be renormalized. However, the principal difference between (3.4) and the GR Lagrangian is that (3.4) already contains arbitrarily high powers of the derivative when perturbatively expanded. This comes from the nonlinearity of the theory, and, as we discussed above, does not contradict the fact that its field equations are second order. It may then be that the higher derivative counter-terms that are needed for the renormalization of a theory from the class (3.4) are of the same general form (3.4). In other words, as we have conjectured, it may be that the class (3.4) is closed under the renormalization.

We then appealed to the ideas of asymptotic safety and gave a possible scenario for the UV-fixed point in this class of theories. Thus, the sought UV-fixed point that controls the dynamics of the theory at ultra-high energies may be the topological fixed point, corresponding to a theory without any propagating degrees of freedom. However, whatever the UV-fixed point may be, if the idea of the closedness of (3.4) under the renormalization is correct, the RG flow (4.12) has a potential of providing a very concrete realization of the asymptotic safety ideas. The RG flow also has a potential to explain why the dimensionless parameter in the Lagrangian of GR (3.28) is so large—which is a rephrasal of the infamous cosmological constant problem.

While these possibilities are intriguing, many questions remain open and much more work is needed to see if the ideas described here can be realized. First and foremost, it is necessary to study how (3.4) gets renormalized in the quantum theory, and see if the expectation of closedness under the renormalization holds. This can be done via the standard textbook methods at least at one-loop level. The required Feynman rules will appear soon. Alternatively, one can use the more sophisticated technology of the background field method and the heat kernel expansion. Work on both of these lines of investigation is in progress. Once it is seen that the class (3.4) is preserved by the renormalization, the RG flow will be calculable and the ideas of asymptotic safety can be probed.

Another very important line of research is that of finding which types of matter can naturally be coupled to gravity in this formalism. As we have discussed above, the only natural coupling in our approach appears to be by enlarging the gauge group from SU(2) that describes pure gravity to a general group G. It would be very interesting to see what type of matter can appear this way. The bosonic case has already been studied by Krasnov (2011*c*) and, in particular, it was found that massless gauge bosons and massive fields of various spins can be described. An interesting version of the spontaneous (gauge) symmetry-breaking scenario was also found in this context. It would be extremely interesting to see if the usual Dirac and Weyl fermions can be coupled to gravity by enlarging the gravitational SU(2) to some super-group.

Another important open problem is to understand the physical implications of the parity-violating character of the present class of theories. The question is already interesting in the context of pure gravity, for it can be, e.g. shown that the propagation of the gravitational waves on any time-dependent (but not conformally flat) homogeneous isotropic background is left–right asymmetric in a general theory from the class (3.4). As it is easy to imagine that departures from GR described by (3.4) could be present in the very early Universe, it is interesting to determine whether there can be any observable effects of the parity asymmetry. This set of questions is also currently under investigation.

References

Ashtekar, A. 1987 New Hamiltonian formulation of general relativity. *Phys. Rev. D* **36**, 1587–1602. (doi:10.1103/PhysRevD.36.1587)

Ashtekar, A. 1991 Lectures on nonperturbative canonical gravity. In *Advanced series in astrophysics and cosmology*, vol. 6, p. 334. Singapore: World Scientific.

Atiyah, M. F., Hitchin, N. J. & Singer, I. M. 1978 Self-duality in four-dimensional Riemannian geometry. *Proc. R. Soc. Lond. A* **362**, 425–461. (doi:10.1098/rspa.1978.0143)

Brans, C. H. 1974 Complex structures and representations of the Einstein equations. *J. Math. Phys.* **15**, 1559. (doi:10.1063/1.1666847)

Bengtsson, I. 1991 The cosmological constants. *Phys. Lett. B* **254**, 55.

Bengtsson, I. 2007 Note on non-metric gravity. *Mod. Phys. Lett. A* **22**, 1643–1649. (doi:10.1142/S0217732307023924)

Berends, F. A. & Giele, W. T. 1988 Recursive calculations for processes with n gluons. *Nucl. Phys. B* **306**, 759–808. (doi:10.1016/0550-3213(88)90442-7)

Capovilla, R. 1992 Generally covariant gauge theories. *Nucl. Phys. B* **373**, 233–246. (doi:10.1016/0550-3213(92)90456-L)

Capovilla, R., Jacobson, T. & Dell, J. 1989 General relativity without the metric. *Phys. Rev. Lett.* **63**, 2325–2328. (doi:10.1103/PhysRevLett.63.2325)

Capovilla, R., Jacobson, T. & Dell, J. 1991 A pure spin connection formulation of gravity. *Class. Quant. Grav.* **8**, 59. (doi:10.1088/0264-9381/8/1/010)

Capovilla, R., Jacobson, T. & Dell, J. 1992 Erratum: a pure spin connection formulation of gravity. *Class. Quant. Grav.* **9**, 1839. (doi:10.1088/0264-9381/9/7/015)

Chakraborty, S. & Peldan, P. 1994 Towards a unification of gravity and Yang–Mills theory. *Phys. Rev. Lett.* **73**, 1195–1198. (doi:10.1103/PhysRevLett.73.1195)

Deser, S. 1970 Self interaction and gauge invariance. *Gen. Rel. Grav.* **1**, 9–18. (doi:10.1007/BF00759198)

Dvali, G., Folkerts, S. & Germani, C. 2011 Physics of trans-Planckian gravity. *Phys. Rev. D* **84**, 024039. (doi:10.1103/PhysRevD.84.024039)

Endlich, S., Nicolis, A., Rattazzi, R. & Wang, J. 2011 The quantum mechanics of perfect fluids. *JHEP* **1104**, 102. (http://arXiv.org/abs/1011.6396)

Feynman, R. P., Morinigo, F. B., Wagner, W. G. & Hatfield, B. 1995 Feynman lectures on gravitation. In *The advanced book program*, p. 232. Reading, MA: Addison-Wesley.

Goroff, M. H. & Sagnotti, A. 1986 The ultraviolet behavior of Einstein gravity. *Nucl. Phys. B* **266**, 709–736. (doi:10.1016/0550-3213(86)90193-8)

Grisaru, M. T., van Nieuwenhuizen, P. & Wu, C. C. 1975 Gravitational born amplitudes and kinematical constraints. *Phys. Rev. D* **12**, 397–403. (doi:10.1103/PhysRevD.12.397)

Hitchin, N. J. ePrint Stable forms and special metrics. e-Print: math/0107101 [math-dg].

Hojman, S. A., Kuchar, K. & Teitelboim, C. 1976 Geometrodynamics regained. *Ann. Phys.* **96**, 88–135. (doi:10.1016/0003-4916(76)90112-3)

Krasnov, K. 2006 Renormalizable non-metric quantum gravity? (http://arXiv.org/abs/hep-th/0611182)

Krasnov, K. 2008 On deformations of Ashtekar's constraint algebra. *Phys. Rev. Lett.* **100**, 081102. (doi:10.1103/PhysRevLett.100.081102)

Krasnov, K. 2009 Plebanski formulation of general relativity: a practical introduction. *Gen. Rel. Grav.* **43**, 1–15. (doi:10.1007/s10714-010-1061-x)

Krasnov, K. 2010 Effective metric Lagrangians from an underlying theory with two propagating degrees of freedom. *Phys. Rev. D* **81**, 084026. (doi:10.1103/PhysRevD.81.084026)

Krasnov, K. 2011*a* Gravity as a diffeomorphism invariant gauge theory. *Phys. Rev. D* **84**, 024034. (doi:10.1103/PhysRevD.84.024034)

Krasnov, K. 2011*b* Pure connection action principle for general relativity. *Phys. Rev. Lett.* **106**, 251103. (doi:10.1103/PhysRevLett.106.251103)

Krasnov, K. 2011*c* Spontaneous symmetry breaking and gravity. (http://arXiv.org/abs/1112.5097)

Krasnov, K. & Shtanov, Y. 2008 Non-metric gravity. II. Spherically symmetric solution, missing mass and redshifts of quasars. *Class. Quant. Grav.* **25**, 025002. (doi:10.1088/0264-9381/25/2/025002)

Krasnov, K. & Shtanov, Y. 2009 Halos of modified gravity. *Int. J. Mod. Phys. D* **17**, 2555. (http://arXiv.org/abs/0805.2668)

MacDowell, S. W. & Mansouri, F. 1977 Unified geometric theory of gravity and supergravity. *Phys. Rev. Lett.* **38**, 739–742. [Erratum. **38**, 1376 (1977)]. (doi:10.1103/PhysRevLett.38.739)

Padmanabhan, T. 2008 From gravitons to gravity: myths and reality. *Int. J. Mod. Phys. D* **17**, 367–398. (doi:10.1142/S0218271808012085)

Penrose, R. & Rindler, W. 1984 *Spinors and space–time.* 1. *Two spinor calculus and relativistic fields.* Cambridge monographs on mathematical physics, p. 458. Cambridge, UK: Cambridge University Press.

Plebanski, J. F. 1977 On the separation of Einsteinian substructures. *J. Math. Phys.* **18**, 2511. (doi:10.1063/1.523215)

't Hooft, G. 1976 Computation of the quantum effects due to a four-dimensional pseudoparticle. *Phys. Rev. D* **14**, 3432–3450. [Erratum. **18**, 2199 (1978)]. (doi:10.1103/PhysRevD.14.3432)

Torres-Gomez, A. & Krasnov, K. 2010 Gravity Yang–Mills–Higgs unification by enlarging the gauge group. *Phys. Rev. D* **81**, 085003. (doi:10.1103/PhysRevD.81.085003)

Torres-Gomez, A., Krasnov, K. & Scarinci, C. 2010 A unified theory of non-linear electrodynamics and gravity. *Phys. Rev. D* **83**, 025023.

Wald, R. M. 1986 Spin-2 fields and general covariance. *Phys. Rev. D* **33**, 3613–3625. (doi:10.1103/PhysRevD.33.3613)

Weinberg, S. 2009 Effective field theory, past and future. *PoS C* **D09**, 001. (http://arXiv.org/abs/0908.1964)

Urbantke, H. 1984 On integrability properties of SU(2) Yang–Mills fields. I. Infinitesimal part. *J. Math. Phys.* **25**, 2321. (doi:10.1063/1.526402)